remote sensing

MDPI

Article

DCN-Based Spatial Features for Improving Parcel-Based Crop Classification Using High-Resolution Optical Images and Multi-Temporal SAR Data

Ya'nan Zhou [1,*], Jiancheng Luo [2,3], Li Feng [1] and Xiaocheng Zhou [4]

1 School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China
2 Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China
3 University of Chinese Academy of Sciences, Beijing 100049, China
4 Key Laboratory of Spatial Data Mining & Information Sharing of Ministry of Education, Fuzhou University, Fuzhou 350116, China
* Correspondence: zhouyn@hhu.edu.cn; Tel.: +86-176-2590-8703

Received: 22 May 2019; Accepted: 5 July 2019; Published: 8 July 2019

Abstract: Spatial features retrieved from satellite data play an important role for improving crop classification. In this study, we proposed a deep-learning-based time-series analysis method to extract and organize spatial features to improve parcel-based crop classification using high-resolution optical images and multi-temporal synthetic aperture radar (SAR) data. Central to this method is the use of multiple deep convolutional networks (DCNs) to extract spatial features and to use the long short-term memory (LSTM) network to organize spatial features. First, a precise farmland parcel map was delineated from optical images. Second, hundreds of spatial features were retrieved using multiple DCNs from preprocessed SAR images and overlaid onto the parcel map to construct multivariate time-series of crop growth for parcels. Third, LSTM-based network structures for organizing these time-series features were constructed to produce a final parcel-based classification map. The method was applied to a dataset of high-resolution ZY-3 optical images and multi-temporal Sentinel-1A SAR data to classify crop types in the Hunan Province of China. The classification results, showing an improvement of greater than 5.0% in overall accuracy relative to methods without spatial features, demonstrated the effectiveness of the proposed method in extracting and organizing spatial features for improving parcel-based crop classification.

Keywords: Sentienl-1 SAR; deep learning; spatial texture feature; time-series analysis; crop classification

1. Introduction

Remote sensing techniques have long been an important means for agricultural monitoring, with their ability to quickly and efficiently collect information about spatial-temporal variability of farmlands and crops [1,2]. Crop-type classification is an essential issue for agricultural monitoring, since it is basic for many applications of precision agriculture (such as crop acreage and yield estimations) [3,4]. While, remote sensing applications in agriculture have traditionally focused on the use of optical satellite data such as Landsat and Gaofen-1 (GF1, imaging satellite of China) images [1,5,6]. However, due to cloud and haze interference (especially in cloudy and rainy regions), optical images are not always available at phenological stages important for crop monitoring, which results in inadequate performance of crop classification. These constraints seriously impede the use of optical images for operational crop mapping [7]. Unlike passive visible and infrared wavelengths which are sensitive to cloud and light,

active synthetic aperture radar (SAR) is particularly attractive for crop classification because of its characteristics of all-weather, all-day imaging capability [7,8].

Based on the advantages of SAR data over optical images, many studies using (multi-temporal) SAR data have obtained great success in agricultural applications (e.g., crop classification and extracting phenological information) in cloudy and rainy regions [9]. However, SAR data have some inherent weaknesses (such as more speckles and fewer bands). They affect the performance of crop-related applications, making it hard to obtain similar (equivalent) accuracies as optical images. Besides (1) combing multi-configuration (e.g., imaging modes, microwave frequencies, incidence angles and polarization modes) SAR data [8,10,11] and (2) constructing a simple and effective measure (similar to the normalized difference vegetation index in optical images) of crop growth [2,4], extracting more and effective spatial (texture) features from SAR data is also a popular research field for SAR-based crop classification, because SAR data measure the interactions between microwave and vegetation and soil, and the spatial structure of vegetation, representing more abundant spatial information.

Many studies focused on spatial feature analysis of SAR data to improve the accuracy of SAR image interpretation [4,7]. Spatial (texture) feature is manifested due to the variations of the measured intensity (e.g., gray level for optical images and scattering coefficient for SAR data) at multiple spatial scales much larger than pixel sizes, which may visually reveal homogeneous or inhomogeneous regions in images [12] and improve the subsequent classification. Many spatial features exist across various scales and should be measured reasonably and effectively. The widely used techniques in previous studies to extract spatial (texture) information involve the Gabor filters, gray level co-occurrence matrix (GLCM), local binary patterns, discrete wavelet transform, and Markov random field (MRF). The most common statistical method is based on GLCM, which characterizes spatial features by calculating how often pairs of pixels with specific values and in a specified spatial relationship occur in an image. The GLCM-based spatial features (e.g., homogeneity, contrast, entropy, and angular second moment) were presented and explored for cocoa agroforest delineation [13], plastic-mulched fields identification [14], early crop type identification [15], winter wheat mapping [4], and crop classification [7]. Qualitative comparative evaluation of the usefulness of GLCM-based spatial features for crop classification demonstrates that spatial textures can improve classification performance, and GLCM-based approaches achieve better results than those of other spatial texture discrimination methods [16]. Further studies on spatial feature calculation parameters were also carried out to find the appropriate scale (moving window size), orientation, displacement, and feature combination for special applications [17,18].

Although encouraging results have been made, the traditional spatial features (e.g., GLCM-based features) for crop classification, face big problems in efficiency and adaptability. Traditional spatial features are often hand designed with domain knowledge at special scales for special applications, which makes limited contributions to improve crop classification [19]. Recently, with the vigorous development of deep learning technologies, deep convolutional networks (DCNs) have been used for learning and extracting multi-scale spatial features in remote sensing applications. By automatically learning hierarchies of spatial features from massive training data [19], DCNs have obtained promising results in single-image-based applications, including satellite image classification [20,21], land-cover classification [22], and crop classification [23,24]. Further studies analyzed the parameter sensitivity of DCNs (e.g., receptive field, convolutional kernel size, depth and number of features) for various application scenarios [25,26]. In the field of remote-sensing time-series classification, a few studies [24,27,28] tried to introduce hybrid frameworks (e.g., the ConvLSTM network [28] and the FCN-LSTM network [24]) of DCNs and recurrent neural networks (RNNs) [29–31] to capture the spatial–temporal features of multi-temporal satellite data. However, these frameworks are pixel-based and hard to transfer to parcel-based crop classification. And previous studies employed these DCN-based spatial features in a black box model, without delving into which features at which scales were more important, how to organize those features in recurrent neural networks, and how to achieve the optimal performance for time-series crop classification.

The objective of this article is to use DCNs to extract multi-scale spatial features to improve parcel-based time-series crop classification using high-resolution optical images and multi-temporal SAR data in cloudy and rainy southern China. We implemented this method through the combined use of DCN-based spatial features and an LSTM-based network structure. Compared to previous approaches, this study combined two principal contributions. The first contribution was employing multiple DCNs to learn hundreds of multi-scale spatial features from SAR data to construct multivariable time series at the parcel level. The second one was designing LSTM-based network structures to organize these features for better classification.

The proposed method is further discussed and validated through parcel-based time-series crop classification on a dataset of ZY-3 (Resources Satellite Three of China) multispectral images and multi-temporal Sentinel-1 SAR data quantitatively, and the results are compared to those without spatial features. Furthermore, using an optimal experimental setting, we achieved the best crop classification.

2. Methods

The DCN-based spatial features were employed to improve time-series analysis for parcel-based crop classification using multi-temporal SAR data and high-resolution optical images, as illustrated in Figure 1. The method involved three main steps: (1) DCN-based spatial features, (2) fine-scale farmland parcel maps, and (3) feature organization and combination for LSTM-based classification.

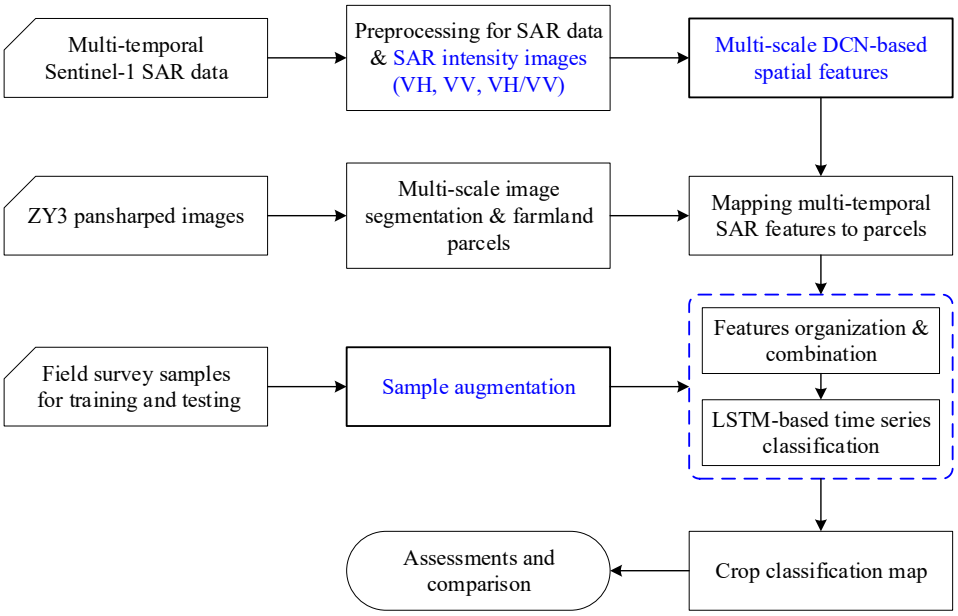


Figure 1. Flowchart of deep convolutional network (DCN)-based spatial features for improving time-series crop classification.

Before the main process, data preprocessing was conducted, which involved the pan-sharpening and mosaic of optical images, preprocessing of multi-temporal SAR data, and geographic registration of experimental data (including SAR data, optical images, and survey samples). First, multi-temporal Sentinel-1A SAR data were first processed to produce intensity images with VH, VV, and the ratio of VH and VV (VH/VV) bands. Then, pre-trained DCNs were applied on intensity images to learn and extract multi-scale spatial features to generate feature images with hundreds of bands. Second, high-resolution optical images were first automatically segmented. Then, on the segmentation map, farmland parcels were identified and simplified (on their boundaries) to produce farmland parcel maps. Third, the multi-temporal SAR feature images were first overlaid onto the parcel map to construct parcel-based time series. Then, time-series features (including VH, VV intensities, and DCN-based spatial features) were organized and combined in an LSTM-based classifier to produce crop classification maps.

2.1. Study Area and Dataset

The study area is in the northern region of Hunan Province, China, with central coordinates of 111°50′E and 29°43′N (Figure 2). The study area, covering a total area of 1210 km^2, contains hilly areas in the north, flatlands in the midwest, and wetlands in the east. It is characterized by a subtropical humid monsoon climate, with an annual average temperature of 16.7 °C, and an annual average rainfall of 1200 mm to 1900 mm. The climatic conditions are suitable for rice, cotton, and rape crop growth. According to the field survey, sequential crop types in a parcel in a year can be classified into (single season) rice, double (season) rice, rice–rape, rape–rice, rape–rice–rape, rice–cotton, and other crop configurations, which are also crop types for classification in this study.

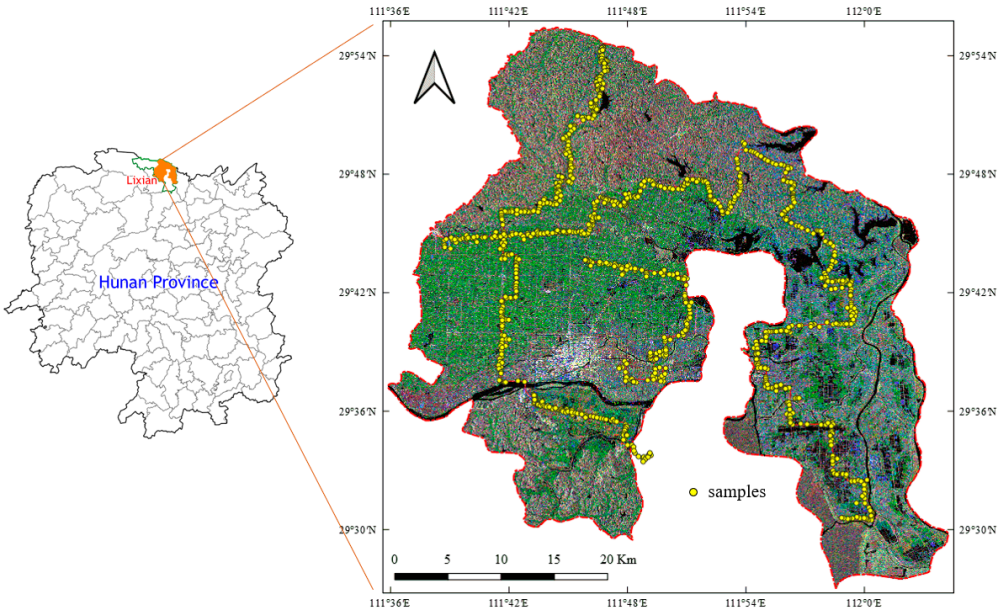


Figure 2. The study area in northern Hunan Province, China. The right-hand section provides an overview of the SAR data (R: 23 May 2017 VH polarization, G: 26 October 2017 VH polarization, B: 3 August 2017 VV polarization) and roadmaps and sample spots of the field survey.

High-resolution ZY3 optical images were used to produce fine-scale parcel maps in this study. The ZY3 image comprises one panchromatic band with a spatial resolution of 2.1 m and four multi-spectral bands (blue, green, red, near-infrared) with spatial resolutions of 5.8 m. Due to their narrow swath of 50 km, two ZY3 images acquired in May 2017 and two ZY3 images acquired in October 2017 (with close acquisition times) were collected to totally cover study areas.

Sentinel-1A SAR data with the interferometric wide swath (with a swath of 250 km and a spatial resolution of 5 m × 20 m) image mode were employed to construct time-series data. They were comprised of a VH polarization and a VV polarization in C band. In this study, we collected all 31 available Sentinel-1A data (distributed as a single-look complex (SLC) product, with Path of 11 and Frame of 94) from 30 December 2016 to 6 January 2018 (with a revisit period of 12 days) from the European Space Agency.

For supervised crop classification and accuracy assessments, three repetitive field surveys were conducted in May, August, and November of 2017. For each parcel, its crop type was recorded three times so that the three surveys fully recorded all potential crop types and rotations for every farmland parcel in one year. Samples from more than 1000 farmland parcels were collected while recording their geographic positions (using a handheld GPS locator with a precise point positioning precision level of 3.0 m, in the WGS84 geographic coordinate system) and crop types through three surveys. The survey roadmap and sample spots are presented in Figure 2. To facilitate the field surveys, samples were distributed along roads, considering topography, geomorphology, hydrology, population, and

economy. Although samples along roads may cause potential decrease in classification performance, we used the same samples in contrast experiments to test the proposed method.

2.2. Data Processing

For the ZY3 image, the panchromatic band and the multi-spectral bands were fused together (using the Gram–Schmidt spectral sharpening algorithm) to produce a pansharpened image with a spatial resolution of 2.1 m. Then, the four images were geometrically registered and mosaiced.

The Sentinel-1A SLC products were processed. Firstly, a thermal noise removal, radiometric correction, and multilooking operation with a looking number of 3 were applied on Sentinel-1A SLC products to generate intensity images with VH bands and VV bands. Secondly, intensity images were filtered by a Lee filter with a 3×3 window to reduce speckle noise. Thirdly, intensity images were geocoded using the shuttle radar topography mission (SRTM) DEM with a 90 m spatial resolution. Their digital number values were converted to a decibel (dB) scale backscatter coefficient. Finally, all intensity images were clipped to cover the study areas. Here, multi-temporal intensity SAR images (with a VH band and a VV band) were obtained. Unfortunately, the Sentinel-1A SLC product on day of year (DOY) 166 (between 154 and 178) was not available from the European Space Agency. So, it was simply restored as the mean feature value of the DOY 154 and DOY 178.

Mosaiced satellite maps (produced from Google Tile Map Service with a spatial resolution of 1.2 m) covering study areas were used as geomatical reference to register all experimental data (including optical images, SAR data, and field survey samples). A polynomial model with 2 degrees and 9 manually selected ground control points (evenly distributed in images) were employed to geometrically correct the ZY3 image to the Google satellite map in ENVI 5.1 software. Clipped sub-SAR data were geometrically corrected following a procedure similar to registration of the ZY3 images. Registration errors were amounted to less than 0.5 pixels at both the ZY3 and SAR spatial resolutions.

2.3. Spatial Features from SAR Data

Besides the VH and VV intensities, spatial features retrieved from SAR data were employed. A traditional spatial feature descriptor (based on GLCM) and three DCN-based spatial feature descriptors (based on the pre-trained VGG16, ResNet50, and DenseNet121 networks in the Keras framework) were utilized to produce spatial feature images in this study.

Since there are only two intensity bands in processed SAR images, we calculated a ratio band (VH/VV) and stacked it into SAR image as the third band to construct a three-band image for learning DCN-based features. The pixel value of the ratio band was defined as the ratio of the VH intensity to the VV intensity in the decibel (dB) scale in the same location. Then, we stretched the ratio values into the 8 bit scale (0, 255) using the maximum value and the minimum value calculated from all multi-temporal ratio bands; similar processes were conducted on the VH and VV band.

2.3.1. GLCM-Based Features

The GLCM was employed to extract traditional features for crop classification. Based on the Haralick analysis, the GLCM-based features were extracted from the three-band SAR images (for experimental contrast with DCN-based features), dependent upon all directions and all image bands, namely, homogeneity, angle second moment, contrast, correlation, dissimilarity, entropy, mean, and standard deviation.

2.3.2. DCN-Based Features

There are many famous DCNs for object detection and instance segmentation in camera images and medical images. Based on open-access datasets (such as COCO and ImageNet) [32,33], several pre-trained DCNs have been released. They can be customized and fine-tuned (removing the last prediction layer and adding a dense layer) for special applications without enough training data [34–37]. Unfortunately, there are not enough SAR sample data available for training new DCNs, or pre-trained

networks available for SAR-based applications. In this study, we tried to employ three released pre-trained DCNs (VGG16, ResNet50, and DenseNet121) to learn SAR-based spatial features to improve crop classification.

There are a great many layers with various depths in DCN architectures. Which layers could produce more effective spatial features? Two principles were employed to guide which layers (spatial features) were selected: (1) the selected layer (feature) should combine as much as possible spatial information at its depth; and (2) the selected layer should stand before a "*pool*" operation to reduce information loss caused by resolution reduction in the "*pool*" operation.

Visual geometry group model (VGG) is a famous convolutional neural network [38]. It is trained on more than a million images belonging to 1000 classes and achieves top 5 test accuracies in the ImageNet database. As a member of the VGG family, the VGG16 architecture consists of 16 convolutional layers. As a result, the network has learned rich feature representations for a wide range of images. Following the principles for selecting layers, we selected the "*block1_conv2*" (feature image of layer 1, resulted from 2 convolution operations), "*block2_conv2*", "*block3_conv3*", "*block4_conv3*", and "*block5_conv3*" layers with their default names from the pre-trained VGG16 model in the Keras framework. Their details are presented in Table 1.

Table 1. Layers and their details selected from the VGG16 model.

Layer Name	Spatial Resolution	Band Number	Label
block1_conv1, block1_conv2, block1_pool	×2	64	V1
block2_conv1, block2_conv2, block2_pool	×4	128	V2
block3_conv1, block3_conv2, block3_conv3, block3_pool	×8	256	V3
block4_conv1, block4_conv2, block4_conv3, block4_pool	×16	512	V4
block5_conv1, block5_conv2, block5_conv3, block5_pool	×32	512	V5

Unlike traditional sequential network architectures (e.g., VGG), residual neural network (ResNet) [39] introduces a novel architecture of skip connections with the hypothesis that the deeper layers should be able to learn something as equal as shallower layers. As one variant of ResNet, ResNet50 with 50 layers was employed to learn very deep spatial features from SAR data in this study. Following the principles for selecting features, five layers of features with their default names were selected from the pre-trained ResNet50 model in the Keras framework, as illustrated in Table 2.

Table 2. Layers and their details selected from the ResNet50 model.

Layer Name	Spatial Resolution (times)	Band Number	Label
bn_conv1	×2	64	R1
add_3	×4	256	R2
add_7	×8	512	R3
add_13	×16	1024	R4
add_16	×32	2048	R5

Dense convolutional network (DenseNet) was proposed with a novel architecture [40] that further exploits the effects of shortcut connections(connecting all layers directly with each other). In this novel architecture, the input of each layer consists of the feature maps of all earlier layers, and its output is passed to each subsequent layer. The feature maps are aggregated with depth concatenation. Using this architecture, DenseNet can aggregate more information from all earlier layers to improve its performance, which makes it superior to previous networks. DenseNet121 is a pretrained (implementation) model with 121 layers, which has been trained on a subset of the ImageNet database. Following the principles for selecting features, five layers of features with their default names were selected from the pre-trained DenseNet121 model in the Keras framework, as illustrated in Table 3.

Table 3. Layers and their details selected from the DenseNet121 model.

Layer Name	Spatial Resolution (times)	Band Number	Label
convl/relu	×2	64	D1
pool2_conv	×4	128	D2
pool3_conv	×8	256	D3
pool4_conv	×16	512	D4
bn	×32	1024	D5

2.4. Plot-Based Time Series Construction

2.4.1. Farmland Parcels Extraction

Farmland parcels were extracted from the ZY3 pansharpened images, through the combined use of an adaptive multi-scale segmentation method [41], automatic identification, and manual correction. First, the segmentation algorithm with default parameter settings was applied on the ZY3 image to produce a segmentation map. Second, a two-class (one was parcel type; the other was non- parcel type) support vector machine classifier (using spectral features of the ZY3 image) available in the scikit-learn package was applied to identify farmland parcels from the segmentation map. Then, interpretation experts checked and corrected object attributes to produce a farmland parcel map. Third, the boundaries of farmland parcels were further corrected and simplified to follow the edges of ground objects in the ZY3 image. The final farmland parcel map is presented in Figure 3, in which each parcel was used as a basic element for time-series analysis. There were approximately 390,000 farmland parcels in the study area.

Figure 3. Farmland parcels and field survey samples for crop classification. Green polygons denote parcels, yellow points denote samples, and the red line denotes the roadmap used when sampling.

2.4.2. Feature Mapping from SAR Data to Parcels

The generated spatial feature images were geometrically overlaid onto the farmland parcel map to construct a time series for the parcels. First, the feature image was interpolated (using a bilinear method) into a size with the same width and height as the Sentinel-1A SAR images. Second, in spatial feature images, pixels within a farmland parcel were searched. Third, the mean feature value of these pixels of a feature band was assigned to the parcel as a separate spatial feature attribute. Four, multiple attributes from the same feature band of multi-temporal SAR data formed a time-series curve. Finally, hundreds of SAR feature curves (one spatial feature to one time-series curve) were constructed for each parcel.

The SAR data are affected little from clouds and shadows, but data missing occasionally occurs. For example, the Sentinel-1A SLC product in DOY 166 is not available from the European Space Agency. In this study, this type of missing data was considered as a system error for the time-series analysis and was simply restored and assigned the mean feature value of the previous (DOY 154) and subsequent time (DOY 178).

2.5. LSTM-Based Time-Series Classification

2.5.1. Sample Augmentation for Classification

More samples are needed to train deep learning models [42] than that to train traditional machine learning algorithms such as the support vector machine and random forest [43,44]. However, real-world samples usually reach a limit, and collecting a volume of field samples would require more manpower, financial resources, and time. This constitutes a major disadvantage of deep learning algorithms used in remote sensing applications [32]. In this study, a limited number of field-survey (FS) samples proved insufficient for training the deep learning model. Thus, some pure pixels within parcel polygons were filtered and selected for sample augmentation to enhance the stability and generalization of classifiers [45].

As is shown in Figure 4, an FS sample spot s was first overlaid onto the farmland parcel map while assigning its crop types t (from the three surveys) to the parcel containing s. Then, an internal buffer region b was constructed along the parcel boundary. The parcel polygon p erased the internal buffer b to create a filtering region f. Finally, each pixel (with its pixel features) within the filtering region f was assigned the crop types t and was independently utilized as a sample. As a result, many more pixel samples (than parcel samples) with hundreds of time-series curves and crop types were obtained for the LSTM-based classification. Using sample augmentation, approximately 5200 pixel samples were obtained for classification in this study.

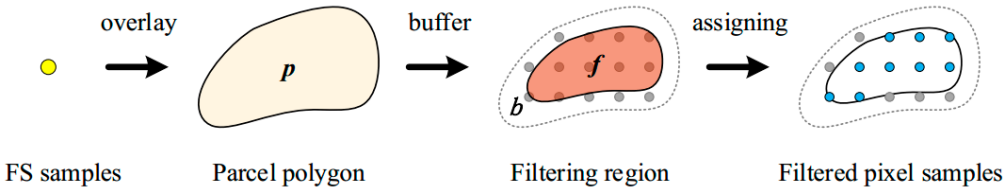


Figure 4. Sample augmentation for the LSTM-based classification.

In experiments, a k-fold cross validation was employed to divide these pixel samples for training and validating classifiers. The speckle filtering applied on the SAR images would introduce some correlation between pixels in the same parcel, which would result in overfitting classifiers and poor performance of sample testing when pixel samples from the same parcel fall into different folds. So, the division of folds was conducted at the parcel level, rather than at the pixel level so that pixels from the same parcel would not belong to different folds.

2.5.2. Structures for Organizing Spatial Features

The LSTM models built based on the Keras framework were employed for time-series crop classification. Thanks to these DCNs (VGG16, ResNet50, and DenseNet121), hundreds of time-series features (curves) at various depths were learned and extracted. But, how to organize and combine these time-series features into the LSTM model? A common practice is to concatenate all these features to feed them into LSTM networks. This structure of concatenation first and LSTM later (referred to as C–L structure) is illustrated in Figure 5. Since they were at various depths and had different value ranges, the DCN-based time-series features were first normalized. Then, all normalized features were concatenated and fed to a stack LSTM layer. Finally, a dense layer and a softmax layer (used for multi-classification problems in neural networks) were added to produce the final crop types.

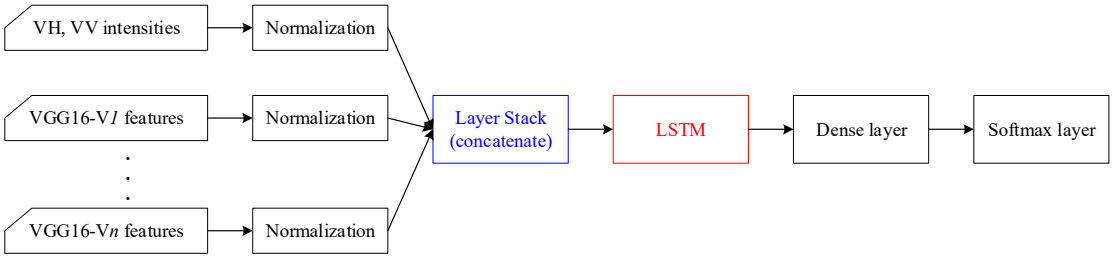


Figure 5. The C–L (concatenation first and LSTM later) structure for organizing and combing time-series features.

An alternative structure was designed to further improve time-series classification in this study. Multiple stack LSTM layers (one feature set to one stack LSTM layer) were first employed to learn and extract time-series features. Then their outputs were normalized and concatenated and fed to a dense layer and a softmax layer to produce the final crop types. This structure of LSTM first and concatenation later (referred to as L–C structure) is illustrated in Figure 6. Each group of features from a special depth of a special DCN were considered as a feature set.

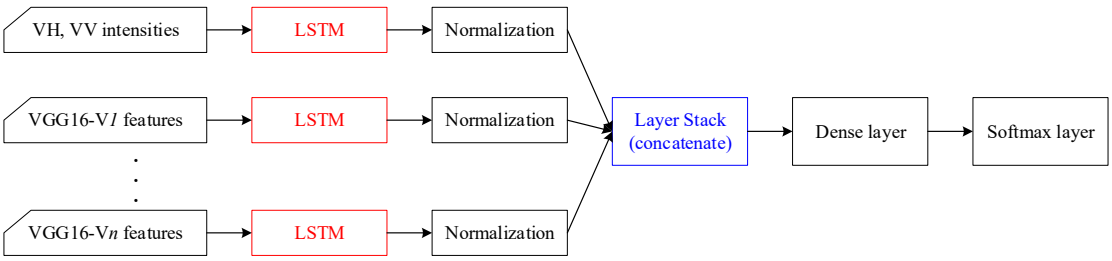


Figure 6. The L–C (LSTM first and concatenation later) structure for organizing and combing time-series features.

In the normalization step, time-series curves were normalized into a range (0, 1.0), using the min-max normalization method, with the minimum and maximum values from all time-series curves in the corresponding feature set. In the LSTM step, a stacked LSTM model with four LSTM layers (the number of hidden neurons is eight times that of the time-series curves) was stacked to transfer raw time-series curves into high-level features. The input shape for the two structures was (sample_number, time_step, feature_number), where sample_number, time_step, and feature_number were the number of training or testing samples, the number of steps in the time-series curves, and the number of extracted features, respectively. In this study, two structures were compared in training time and classification accuracy. We further discussed whether and how to split feature sets into more subgroups in the L–C structure for better performance.

2.6. Performance Evaluation

Based on the confusion matrix derived by comparing classified results to the test samples parcel by parcel, user's accuracy (UA), producer's accuracy (PA), overall accuracy (OA), kappa coefficient [46], and F1 scores [47] were retrieved to evaluate the accuracy of crop classification. The OA was computed by dividing all correctly classified parcels by the entire validation dataset. The kappa was computed to determine whether the values in an error matrix were significantly better than the values of a random assignment. Additionally, $F1 = 2 \times UA \times PA/(UA + PA)$, the harmonic mean of the PA and UA, was more meaningful than the kappa coefficient and the OA for a special class. Larger UA, PA, OA, kappa, and F1 measures denote better results and vice versa.

3. Experiments and Discussion

Using high-resolution ZY3 optical images and multi-temporal Sentinel-1A SAR data, experiments were carried out to test the effectiveness and efficiency of the proposed method. Accuracy assessments were conducted on the parcel-based classification results to discuss whether DCN-based spatial features can improve time-series crop classification, and how to organize these spatial features from multiple DCNs and multiple depths to achieve the best classification performance.

3.1. Evaluation and Discussion

The DCN-based spatial features were employed to improve SAR-based time-series crop classification. To achieve the best classification performance, we conducted a series of experiments on whether DCN-based features were useful, which features at which depths were more effective, and which structures are better to to organize these features.

3.1.1. DCN-Based Features versus GLCM-Based Features

Compared to the traditional spatial feature extractor GLCM, DCNs can learn hundreds of features. They represent more complete spatial information to benefit crop classification. In this experiment, the L–C structure was employed to produce four crop classification maps, using the GLCM-based, VGG16-V1-based, ResNet50-R1-based, and DenseNet121-D1-based spatial features with input shapes of (64 features, 32 steps), respectively. Their classification accuracies (OA and kappa scores for the whole map) are presented in Table 4.

Table 4. Comparison of OAs and kappa scores conducted using the GLCM-, VGG16-V1-, ResNet50-R1-, and DenseNet121-D1-based time-series features.

	VH/VV (baseline)	**VH/VV GLCM**	**VH/VV VGG16-V1**	**VH/VV ResNet50-R1**	**VH/VV DenseNet121-D1**
OA (%)	83.32	84.49	86.93	87.61	**87.87**
kappa	80.28	80.74	81.24	82.13	**82.96**

It is clear that combining spatial features can substantially improve time-series classification (with increases of 1%~5% in OA), compared to classification using VH and VV intensities. This result is consistent with previous studies [4,16]. In detail, improvements using GLCM-, VGG16-V1-, ResNet50-R1-, and DenseNet121-D1-based spatial features were approximately 1.1%, 3.6%, 4.3%, and 4.5% in OA, respectively. Classification with DCN-based spatial features achieved much higher accuracies than that with the traditional GLCM-based features. There were two advantages of DCN-based features over GLCM-based features: (1) trained using million images, the pre-trained DCNs could extract more effective spatial information than the manual designed GLCM; and (2) hundreds of spatial features retrieved from DCNs could represent more complete and more comprehensive spatial information than GLCM-based features.

It was also noted that there were small differences in OA values among the VGG16-V1-, ResNet50-R1-, and DenseNet121-D1-based classifications. DenseNet121-D1-based features achieved the best performance of 87.87% in OA, followed by ResNet50-R1-based features with 87.61% in OA, and VGG16-V1-based classification the worst with 86.93% in OA. Although differences in OAs were very small, they also implied the capability of DCNs for spatial feature learning. On the one hand, DenseNet121 had deeper layers to extract higher-level features than ResNet50 and VGG16. On the other hand, DenseNet121 could aggregate more information from all earlier layers (using its shortcut connections) to enhance spatial features, compared with ResNet50 and VGG16.

3.1.2. Depth of DCN-Based Features

Compared to the traditional GLCM-based feature, DCN-based spatial features were deeper. They represent multi-scale (from the pixel scale to global scale) spatial information to improve crop classification, while deeper features also indicate coarser spatial resolution for feature images. Coarser-resolution feature images would weaken the attribute differences of adjacent parcels, and further degrade accuracies of parcel-based crop classifications. In this study, we tried to discuss DCN-based features at which depths would improve time-series classification and whether deeper features would benefit crop classification.

The VGG16-based spatial features were used to conduct three groups of classification experiments. The first group involved five time-series classifications using V1 (64 features, 32 steps), V2 (128 features, 32 steps), V3 (256 features, 32 steps), V4 (512 features, 32 steps), and V5 (512 features, 32 steps) feature sets, respectively. The second group involved five classifications using V1, V1+V2, V1+V2+V3, V1+V2+V3+V4, and V1+V2+V3+V4+V5 feature sets, respectively. The third group involved six time-series classifications using VH/VV, VH/VV+V1, VH/VV+V1+V2, VH/VV+V1+V2+V3, VH/VV+V1+V2+V3+V4, and VH/VV+V1+V2+V3+V4+V5 feature sets, respectively. Here, "+" indicates organizing spatial features in an L–C structure. Feature sets used are presented in Table 5.

Table 5. Details on the combinations of feature sets.

First Group		Second Group		Third Group	
Label	**Feature Sets**	**Label**	**Feature Sets**	**Label**	**Feature Sets**
				M0	VH/VV
V1	V1	**F1**	V1	**M1**	VH/VV+V1
V2	V2	**F2**	V1+V2	**M2**	VH/VV+V1+V2
V3	V3	**F3**	V1+V2+V3	**M3**	VH/VV+V1+V2+V3
V4	V4	**F4**	V1+V2+V3+V4	**M4**	VH/VV+V1+V2+V3+V4
V5	V5	**F5**	V1+V2+V3+V4+V5	**M5**	VH/VV+V1+V2+V3+V4+V5

Table 6 presents the classification accuracies (OA and kappa values) resulting from the experiments using the feature combinations presented in Table 5. In the first group, along with the deepening depth of the feature sets, the classification accuracies reduced sharply, from 81.27% in OA (using the V1 combination) to 53.81% in OA (using the V3 combination) (because overall accuracy in the V3 classification was very low, so we ignored the accuracies of the V4 and V5 classifications). This was expected, because deeper features would reduce the spatial resolution of feature images, which would result in smaller crop differences and classification accuracies. In the second group, classification accuracies stayed at relatively stable levels. Although, the best accuracies (82.84% in OA) were achieved using the F5 combination, it was still worse than that using VH/VV intensities (as illustrated by the M0 combination in the third group). Since the spatial resolution of feature images decreased along with deepening depths, the highest spatial resolution in this group (F1) was two times that of the SAR data, which resulted in the lower performance of this group. Compared to the second group, the third group incorporated the VH/VV intensity features, which markedly improved classification accuracies.

However, improvements in classification accuracies were very little from the M2 combination to the M5 combination, along with more and deeper features combined.

Table 6. Comparison of overall accuracy and kappa score of classifications using feature combinations.

First group		**V1**	**V2**	**V3**	**V4**	**V5**
OA (%)		81.27	71.48	53.81	-	-
kappa		78.82	63.37	40.85	-	-
Second group		**F1**	**F2**	**F3**	**F4**	**F5**
OA (%)		81.27	81.56	82.39	82.31	82.84
kappa		78.82	79.27	78.53	77.92	78.09
Third group	**M0**	**M1**	**M2**	**M3**	**M4**	**M5**
OA (%)	*83.32*	87.61	**88.15**	88.23	88.09	88.31
kappa	*80.28*	82.13	**82.84**	82.96	82.79	82.71

The results suggest that, although DCNs can learn very deep spatial features, it also lowers the quality of the spatial resolution of feature maps. Thus, it is hard to achieve better classification using DCN-based features independently. A combination of the VH/VV intensity images and DCN-based features should be used. Moreover, along with decreasing the spatial resolutions of DCN-based feature images, their effects on classification decreased, thus, the first two depths may be an optimal choice for using DCN-based spatial features.

3.1.3. Better Structure of Classification Network

The DCNs were employed to learn hundreds of spatial features for time-series crop classification. To combine those spatial features, we implemented two network structures in Section 2.4: C–L and L–C. Through two experiments, we checked which structure could preferably organize these hundreds of spatial features to improve classification.

In the first experiment, the C-L and L-C structures were used to organize the VH/VV features (2 features, 32 steps, as one feature set) and the GLCM-based, the first-layer DCN-based spatial features (64 features, 32 steps, as the other feature set), respectively. As described in Section 2.5, in the C–L structure, we first concatenated the VH/VV features and the DCN-based features into a multi-band image, and then fed it into the stack LSTM network to obtain the final classification. While in the L–C structure, we first fed, respectively, the VH/VV features and the DCN-based features into stack LSTM networks, and then concatenated their output for the final classification. Classification accuracies resulting from the C–L and L–C structures are presented in Table 7.

Table 7. Comparisons of the OA and kappa coefficient conducted in the C–L structure and L–C structures.

		VH/VV GLCM	**VH/VV VGG16-V1**	**VH/VV ResNet50-R1**	**VH/VV DenseNet121-D1**
C-L	OA (%)	84.32	85.23	85.27	85.84
	kappa	80.86	81.79	82.45	82.61
L-C	OA (%)	84.49	86.93	87.61	87.87
	kappa	80.74	81.24	82.13	82.96

Overall, the L–C structure achieved better classification accuracies than the C–L structure, with improvements of 0.5%~2.0% in OA. In the C–L structure, all feature sets were concatenated to determine crop types; while in the L–C structure, each feature set was first separately employed to learn time-series features of crop types (in the L step), and then concatenated their features to determine the final crop types (in the C step). From the procedure point of view, the L–C structure is more like an ensemble learning process, which uses multiple classifications from sub-features to obtain better performance than that could be obtained from any of the classifications alone. Similarly, the L–C structure first split

spatial feature sets into multiple sub-feature sets for extracting time-series features for each subset, in the "L" step, and then concatenated these sub-feature sets for the final classification, in the "C" step. Thus, the L–C structure achieved a better performance than the C–L structure.

In the above experiment, the L–C structure achieved better performance. In the L–C structure, could splitting features into more sub-feature sets further improve time-series crop classification? Taking the VGG16-based multi-depth features as input, we implemented five settings (feature combinations) for experiments, as presented in Table 8, where "/" indicates the C–L structure and "+" indicates the L–C structure, and Vn_i *(n = 1, 2, 3, 4, 5; I = 1, 2, 4, 8)* indicates that the ith feature subsets at the nth depth, which would be used as a feature set. Take the S2 setting as an example: each feature set Vn would be split into two subsets ($Vn \rightarrow Vn_1$ and Vn_2), and the VH feature set, VV feature set, and each feature subset would separately feed into the L–C structure.

Table 8. Five feature combinations for organizing features in the L–C structure.

Label	Structure of Network	# Feature Sets
S0	VH/VV/V1/V2/V3/V4/V5	7
S1	VH+VV+V1+V2+V3+V4+V5	7
S2	VH+VV+$V1_1$+$V1_2$+$V2_1$+$V2_2$+$V3_1$+$V3_2$+$V4_1$+$V4_2$+$V5_1$+$V5_2$	12
S4	VH+VV+$V1_1$+$V1_2$+ … +$V1_4$+$V2_1$+$V2_2$+ … +$V2_4$+$V3_1$+$V3_2$+ … +$V3_4$+$V4_1$+$V4_2$+ … +$V4_4$+$V5_1$+$V5_2$+ … +$V5_4$	22
S8	VH+VV+$V1_1$+$V1_2$+ … +$V1_8$+$V2_1$+$V2_2$+ … +$V2_8$+$V3_1$+$V3_2$+ … +$V3_8$+$V4_1$+$V4_2$+ … +$V4_8$+$V5_1$+$V5_2$+ … +$V5_8$	42

Table 9 presents the performance comparison of classifications using these five feature combinations. The S1 feature combination achieved the best classification (with an OA of 88.26%), followed by the S2 combination (with an OA of 87.56%), the S4 combination (with an OA of 86.91%), and S0 combination (with an OA of 85.42%), followed by the S8 combination with the worst performance (with an OA of 84.02%). It was expected that the S0 combination would produce the lower classification accuracies, because it was actually the C–L structure. From the S1 to S8 combinations, the classification accuracies gradually decreased, along with finer splits on the DCN-based spatial features. From the perspective of information theory, all features (no matter the number of features) at the same depth in DCNs are complete enough to represent ground objects, while finer splits would hurt the completeness of the feature combinations for identifying crop types (any feature subsets lacking enough information), resulting in an increasingly worse performance.

Table 9. Performance comparison of the five feature combinations.

	S0	S1	S2	S4	S8
OA (%)	85.42	**88.26**	87.56	86.91	84.02
kappa	80.27	**82.77**	81.43	80.84	80.80
Epoch time(s)	**21**	37	58	113	194
Epochs	49	**41**	61	50	-

Further, we counted how many epochs were used to reach a stable performance (classification accuracy) and how long the epochs were in the five combinations, as illustrated in Table 9 and Figure 7 (because of different hardware environments and relative time was used for epoch time). Along with the increase in feature subsets in the L–C structure (from the S0 to S8), the time consumed in each epoch increased from 21 to 194. While, the S1 combination used 41 epochs to reach its stable performance, less than 49 epochs were used in the S0 combination, 61 epochs in the S2 combination, and 50 epochs in the S3 combination.

Thus, we think that (1) the L–C structure could achieve higher classification accuracies than the C–L structure and (2) that taking all spatial features at a special depth as a feature set in the L–C structure is the optimal organization.

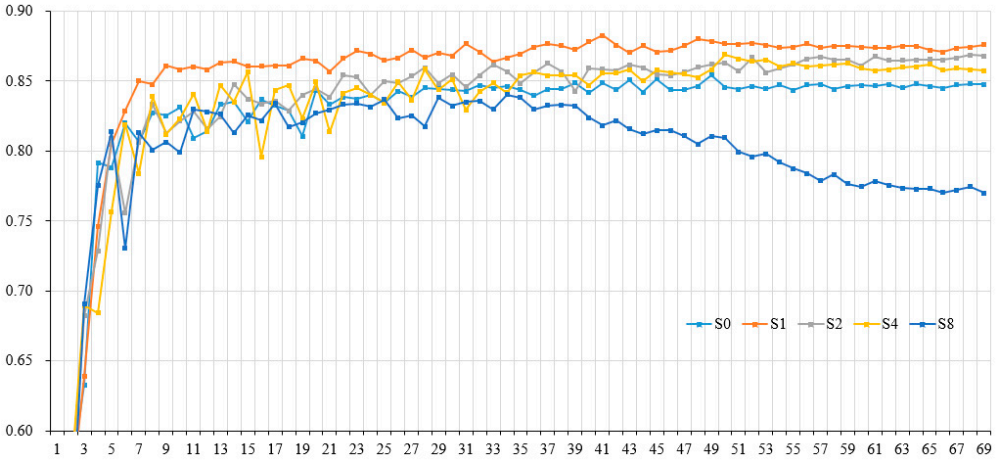
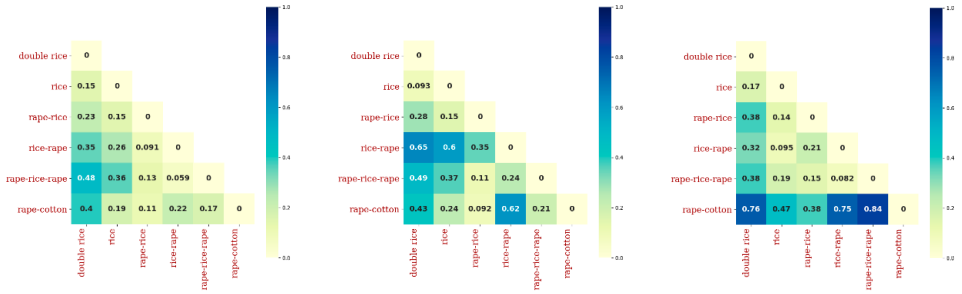


Figure 7. Classification accuracies along with epochs for the five combinations.

3.1.4. Which Features Benefit Which Crops?

On the classification maps produced in Section 3.2, 50 parcels for each crop type were first randomly selected to calculate their mean time-series curves of VGG16-V1, ResNet50-R1, and DenseNet121-D1 features (64 features, 32 steps). Then Fréchet distance [48] was employed to measure the time-series separation between any two crops. The distance matrixes using VGG16-V1, ResNet50-R1, and DenseNet121-D1 features are presented in Figure 8, in which greater values indicate greater separation between the two crops and benefit crop classification.



(**a**) Distance matrix for VGG16-V1 features

(**b**) Distance matrix for ResNet50-R1 features

(**c**) Distance matrix for DenseNet121-D1 features

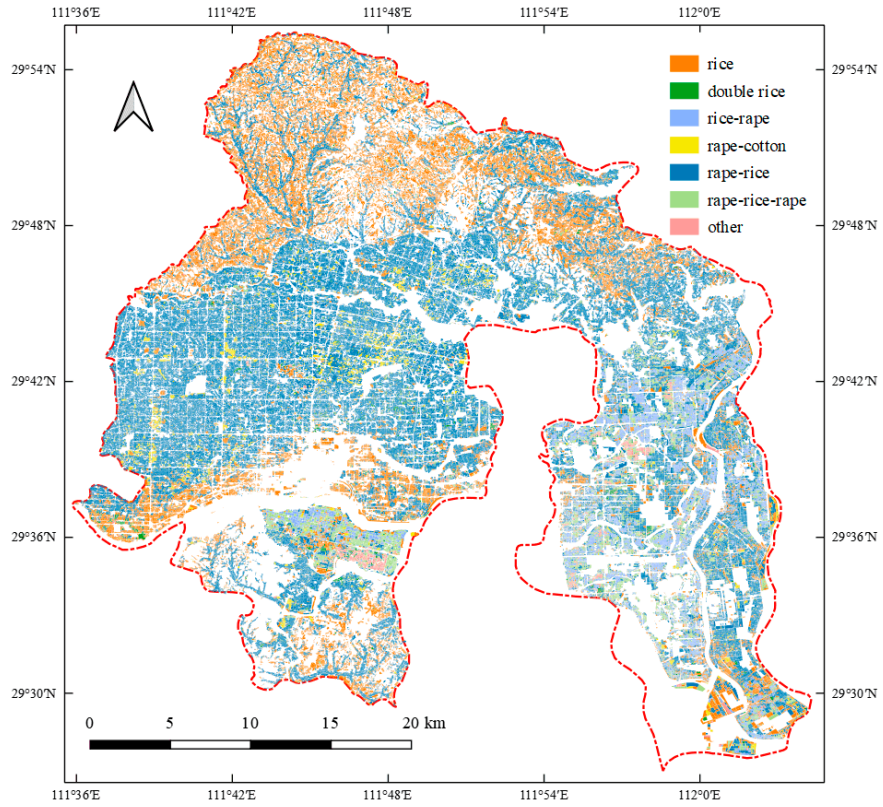
Figure 8. Distance matrixes of crops using VGG16-V1, ResNet50-R1, and DenseNet121-D1 features.

On the whole, the distance matrix of DenseNet121-D1 features had more and greater distance values than the distance matrix for ResNet50-R1 features, and the distance matrix for VGG16-V1 features had the least and smallest distance values. Thus, classifications resulted from DenseNet121-D1 features achieved higher accuracies than that from the ResNet50-R1 features and VGG16-V1 features, which is constant with the results in Tables 4 and 9. In detail, in the distance matrix for VGG16-V1 features, there were more and greater distances between "double rice" crops and other crops. In the distance matrix for ResNet50-R1 features, greater distances mainly appeared between "rice–rape" crops and other crops, between "double rice" crops and other crops. In the distance matrix for DenseNet121-D1 features, distances between "rape–cotton" crops and other crops were far greater, while "rice" crops showed smaller distances with other crops (except for "rape–cotton" crops). These greater distances would benefit from distinguishing different crops, e.g., "double rice", "rice–rape", and "rape–cotton" crops. Conversely, smaller distances between "rape–rice" crops and other crops in the matrix for VGG16-V1 features and ResNet50-R1 features would decrease their classification accuracies.

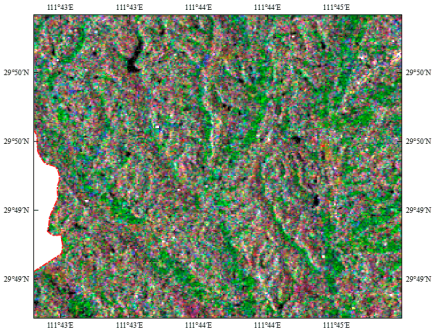
Further, different DCNs have various abilities to represent different crops. Thus, incorporating spatial features from multiple DCNs would improve overall accuracies of time-series crop classification.

3.2. The Optimal Classification Results

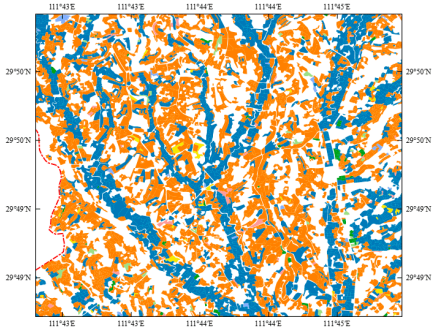
Based on the above discussions, an optimal experiment setting for time-series classification using DCN-based spatial features was constructed to achieve the best performance. In this setting, we employed the L–C structure to organize the VH/VV intensity bands and the VGG16-V1, VGG16-V2, and ResNet50-R1 feature sets (64 features, 32 steps), and ResNet50-R2, DenseNet121-D1, and DenseNet121-D2 feature sets (128 features, 32 steps) to produce the optimal classification map. The optimal classification map with a spatial resolution of 2.1 m and some local details (with corresponding SAR data with the same color combinations as those used in Figure 5) was produced and is presented in Figure 9.



Crop classification map of the study area in Hunan Province.

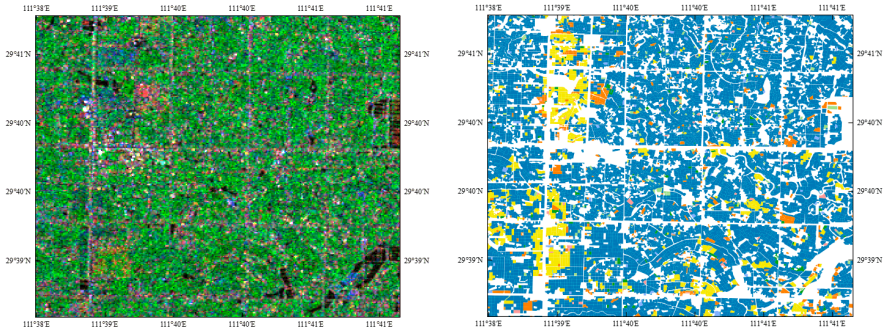


(**a1**) Local details of SAR data in region A.

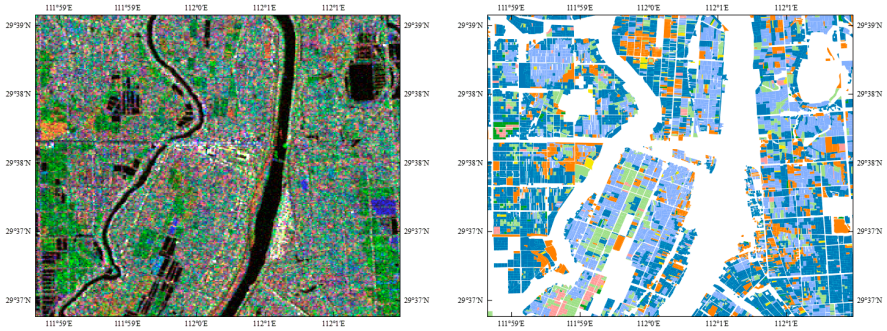


(**a2**) Classification results of region A.

Figure 9. *Cont.*



(**b1**) Local details of SAR data in region B. (**b2**) Classification results of region B.



(**c1**) Local detail of SAR data in region C. (**c2**) Classification results of region C.

Figure 9. Results of time-series crop classification for Hunan Province using the optimal experiment setting.

Classification accuracies of UA, PA, and F1 for each crop and OA and kappa for the whole map are presented in Table 10.

Table 10. Classification accuracies using the optimal experiment setting.

Crop	Rice	Double Rice	Rice–Rape	Rape–Cotton	Rape–Rice	Rape–Rice–Rape	Other
UA (%)	84.82	91.33	92.64	**93.63**	86.72	90.99	83.29
PA (%)	86.31	**91.94**	91.18	91.12	87.67	89.70	82.75
F1	0.8556	0.9163	0.9190	**0.9236**	0.8719	0.9034	0.8302
OA (%)				89.41			
kappa				0.8293			

It is clear that the classification with the optimal experiment setting substantially achieved higher accuracies, compared with the above classifications, with an OA of 89.41% for the whole classification map. In detail, the "double rice", "rice–rape", and "rape–cotton" crops obtained higher OA and kappa values, while the "rice" crop obtained lower values. This is constant with the discussion in Section 3.2, as the "*other*" crop may include a combination of multiple unspecified crops, it achieved the poorest performance of all three classifiers.

4. Conclusions

Multi-scale spatial features retrieved from SAR data are essential to parcel-based time-series crop classification in cloudy and rainy southern China. We proposed a deep-learning-based time-series analysis method for improving parcel-based crop classification, employing multiple DCNs to learn hundreds of multi-scale spatial features and organizing them into LSTM-based time-series classifiers. The method was applied to produce a parcel-based crop classification map from a dataset of ZY3 optical images and Sentinel-1A SAR data for the Hunan Province in China. The optimal classification results, showing an improvement of greater than 5.0% in overall accuracy relative to methods without spatial

features, demonstrated the effectiveness of the proposed method in extracting and organizing spatial features for improving parcel-based crop classification. From further discussions on the effectiveness of DCN-based spatial features, depths of DCN-based features, and organization of DCN-based spatial features, this study concludes: (1) DCN-based spatial features could further improve time-series crop classification relative to traditional GLCM-based features; and (2) the structure of "LSTM first and concatenation later" could further organize these DCN-based features to improve time-series crop classification.

Despite these encouraging results, more work is needed to further explore the use of deep learning for improving time-series crop classification such as (1) collecting more SAR data to train special DCNs or autoencoder models suitable for SAR-based applications; and (2) designing DCN–LSTM-based frameworks to extract and organize thousands of spatial–temporal-spectral features to improve parcel-based crop classification.

Author Contributions: Conceptualization, Y.Z. and J.L.; formal analysis, L.F.; funding acquisition, Y.Z.; investigation, L.F.; methodology, Y.Z.; project administration, X.Z.; resources, J.L. and X.Z.; software, Y.Z.; supervision, J.L. and X.Z.; validation, L.F.; writing—original draft, Y.Z.; writing—review and editing, Y.Z. and L.F.

Funding: This research was funded by the National Natural Science Foundation of China grant number 41631179, the National Key Research and Development Program of China, grant number 2017YFB0503600, the Fundamental Research Funds for the Central Universities, grant number 2019B17114, and the Opening Foundation of Key Lab of Spatial Data Mining & Information Sharing, Ministry of Education (Fuzhou University), grant number 2019LSDMIS04.

Acknowledgments: We thank the anonymous reviewers for their insights and constructive comments to help improve the paper.

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

References

1. Yang, Y.; Huang, Q.; Wu, W.; Luo, J.; Gao, L.; Dong, W.; Wu, T.; Hu, X. Geo-Parcel Based Crop Identification by Integrating High Spatial-Temporal Resolution Imagery from Multi-Source Satellite Data. *Remote Sens.* **2017**, *9*, 1298. [CrossRef]
2. 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]
3. Shao, Y.; Fan, X.; Liu, H.; Xiao, J.; Ross, S.; Brisco, B.; Brown, R.; Staples, G. Rice monitoring and production estimation using multitemporal RADARSAT. *Remote Sens. Environ.* **2001**, *76*, 310–325. [CrossRef]
4. Zhou, T.; Pan, J.; Zhang, P.; Wei, S.; Han, T. Mapping Winter Wheat with Multi-Temporal SAR and Optical Images in an Urban Agricultural Region. *Sensors* **2017**, *17*, 1210. [CrossRef] [PubMed]
5. Gao, F.; Anderson, M.C.; Zhang, X.; Yang, Z.; Alfieri, J.G.; Kustas, W.P.; Mueller, R.; Johnson, D.M.; Prueger, J.H. Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sens. Environ.* **2017**, *188*, 9–25. [CrossRef]
6. Piedelobo, L.; Hernández-López, D.; Ballesteros, R.; Chakhar, A.; Del Pozo, S.; González-Aguilera, D.; Moreno, M.A. Scalable pixel-based crop classification combining Sentinel-2 and Landsat-8 data time series: Case study of the Duero river basin. *Agric. Syst.* **2019**, *171*, 36–50. [CrossRef]
7. Jia, K.; Li, Q.; Tian, Y.; Wu, B.; Zhang, F.; Meng, J. Crop classification using multi-configuration SAR data in the North China Plain. *Int. J. Remote Sens.* **2012**, *33*, 170–183. [CrossRef]
8. 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]
9. Park, S.; Im, J.; Park, S.; Yoo, C.; Han, H.; Rhee, J. Classification and Mapping of Paddy Rice by Combining Landsat and SAR Time Series Data. *Remote Sens.* **2018**, *10*, 447. [CrossRef]
10. Skriver, H. Crop Classification by Multitemporal C-and L-Band Single-and Dual-Polarization and Fully Polarimetric SAR. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 2138–2149. [CrossRef]
11. Zeyada, H.H.; Ezz, M.M.; Nasr, A.H.; Shokr, M.; Harb, H.M. Evaluation of the discrimination capability of full polarimetric SAR data for crop classification. *Int. J. Remote Sens.* **2016**, *37*, 2585–2603. [CrossRef]

12. Kandaswamy, U.; Adjeroh, D.; Lee, M. Efficient texture analysis of SAR imagery. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2075–2083. [CrossRef]
13. Numbisi, F.N.; Van Coillie, F.; De Wulf, R. Delineation of Cocoa Agroforests Using Multiseason Sentinel-1 SAR Images: A Low Grey Level Range Reduces Uncertainties in GLCM Texture-Based Mapping. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 179. [CrossRef]
14. Lu, L.; Tao, Y.; Di, L. Object-Based Plastic-Mulched Landcover Extraction Using Integrated Sentinel-1 and Sentinel-2 Data. *Remote Sens.* **2018**, *10*, 1820. [CrossRef]
15. Inglada, J.; Vincent, A.; Arias, M.; Marais-Sicre, C. Improved Early Crop Type Identification by Joint Use of High Temporal Resolution SAR and Optical Image Time Series. *Remote Sens.* **2016**, *8*, 362. [CrossRef]
16. Reed, T.R.; Dubuf, J.M.H. A review of recent texture segmentation and feature extraction techniques. *CVGIP Image Underst.* **1993**, *57*, 359–372. [CrossRef]
17. Hall-Beyer, M. Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales. *Int. J. Remote Sens.* **2017**, *38*, 1312–1338. [CrossRef]
18. Lan, Z.; Liu, Y. Study on Multi-Scale Window Determination for GLCM Texture Description in High-Resolution Remote Sensing Image Geo-Analysis Supported by GIS and Domain Knowledge. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 175. [CrossRef]
19. Chen, S.; Wang, H.; Xu, F.; Jin, Y.-Q. Target Classification Using the Deep Convolutional Networks for SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1–12. [CrossRef]
20. Lv, X.; Ming, D.; Chen, Y.; Wang, M. Very high resolution Remote sensing image classification with SEEDS-CNN and scale effect analysis for superpixel CNN classification. *Int. J. Remote Sens.* **2019**, *40*, 506–531. [CrossRef]
21. Zhao, W.; Guo, Z.; Yue, J.; Zhang, X.; Luo, L. On combining multiscale deep learning features for the classification of hyperspectral Remote sensing imagery. *Int. J. Remote Sens.* **2015**, *36*, 3368–3379. [CrossRef]
22. Liu, S.; Qi, Z.; Li, X.; Yeh, A.G.-O. Integration of Convolutional Neural Networks and Object-Based Post-Classification Refinement for Land Use and Land Cover Mapping with Optical and SAR Data. *Remote Sens.* **2019**, *11*, 690. [CrossRef]
23. 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]
24. Teimouri, N.; Dyrmann, M.; Jørgensen, R.N. A Novel Spatio-Temporal FCN-LSTM Network for Recognizing Various Crop Types Using Multi-Temporal Radar Images. *Remote Sens.* **2019**, *11*, 990. [CrossRef]
25. Zhao, W.; Du, S. Learning multiscale and deep representations for classifying Remotely sensed imagery. *ISPRS J. Photogramm. Remote Sens.* **2016**, *113*, 155–165. [CrossRef]
26. Romero, A.; Gatta, C.; Camps-Valls, G. Unsupervised Deep Feature Extraction for Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1349–1362. [CrossRef]
27. Liu, Q.; Zhou, F.; Hang, R.; Yuan, X. Bidirectional-Convolutional LSTM Based Spectral-Spatial Feature Learning for Hyperspectral Image Classification. *Remote Sens.* **2017**, *9*, 1330.
28. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. Available online: https://arxiv.org/abs/1506.04214 (accessed on 30 June 2019).
29. Zhong, L.; Hu, L.; Zhou, H. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [CrossRef]
30. Rußwurm, M.; Korner, M. Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 11–19.
31. Ndikumana, E.; Minh, D.H.T.; Baghdadi, N.; Courault, D.; Hossard, L. Deep Recurrent Neural Network for Agricultural Classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sens.* **2018**, *10*, 1217. [CrossRef]
32. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the European Conference on Computer Vision Proceedings of the Computer Vision–ECCV 2014, Zurich, Switzerland, 6–12 September 2014.
33. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2014**, *115*, 211–252. [CrossRef]

34. Han, X.; Zhong, Y.; Cao, L.; Zhang, L. Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification. *Remote Sens.* **2017**, *9*, 848. [CrossRef]
35. Nogueira, K.; Penatti, O.A.; Dos Santos, J.A. Towards better exploiting convolutional neural networks for Remote sensing scene classification. *Pattern Recognit.* **2017**, *61*, 539–556. [CrossRef]
36. Hu, F.; Xia, G.-S.; Hu, J.; Zhang, L. Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. *Remote Sens.* **2015**, *7*, 14680–14707. [CrossRef]
37. Castelluccio, M.; Poggi, G.; Sansone, C.; Verdoliva, L. Land Use Classification in Remote Sensing Images by Convolutional Neural Network. *arXiv* **2015**, arXiv:1508.00092. Available online: https://arxiv.org/abs/1508.00092 (accessed on 30 June 2019).
38. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556. Available online: https://arxiv.org/abs/1409.1556 (accessed on 30 June 2019).
39. 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, 26 June—1 July 2016.
40. Huang, G.; Liu, Z.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
41. Zhou, Y.; Li, J.; Feng, L.; Zhang, X.; Hu, X. Adaptive Scale Selection for Multiscale Segmentation of Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3641–3651. [CrossRef]
42. Zhang, R.; Li, W.; Mo, T. Review of Deep Learning. *arXiv* **2018**, arXiv:1804.01653. Available online: https://arxiv.org/abs/1804.01653 (accessed on 30 June 2019).
43. Zhang, L.; Zhang, L.; Du, B. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [CrossRef]
44. Ball, J.E.; Anderson, D.T.; Chan, C.S. A Comprehensive Survey of Deep Learning in Remote Sensing: Theories, Tools and Challenges for the Community. *J. Appl. Remote Sens.* **2017**, *11*, 042609. [CrossRef]
45. Ding, J.; Chen, B.; Liu, H.; Huang, M. Convolutional Neural Network With Data Augmentation for SAR Target Recognition. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 364–368. [CrossRef]
46. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*; CRC press: Boca Raton, FL, USA, 2008.
47. Sasaki, Y. The truth of the F-measure. *Teach Tutor Mater* **2007**, *1*, 1–5.
48. Ahn, H.K.; Knauer, C.; Scherfenberg, M.; Schlipf, L.; Vigneron, A. Computing the discrete Fréchet distance with imprecise input. *Int. J. Comput. Geom. Appl.* **2012**, *22*, 27–44. [CrossRef]