



Article

An Adversarial Generative Network for Crop Classification from Remote Sensing Timeseries Images

Jingtao Li, Yonglin Shen and Chao Yang *

School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China; JingtaoLi@cug.edu.cn (J.L.); shenyl@cug.edu.cn (Y.S.)

* Correspondence: yangchao@cug.edu.cn

Abstract: Due to the increasing demand for the monitoring of crop conditions and food production, it is a challenging and meaningful task to identify crops from remote sensing images. The state-of-the-art crop classification models are mostly built on supervised classification models such as support vector machines (SVM), convolutional neural networks (CNN), and long- and short-term memory neural networks (LSTM). Meanwhile, as an unsupervised generative model, the adversarial generative network (GAN) is rarely used to complete classification tasks for agricultural applications. In this work, we propose a new method that combines GAN, CNN, and LSTM models to classify crops of corn and soybeans from remote sensing time-series images, in which GAN's discriminator was used as the final classifier. The method is feasible on the condition that the training samples are small, and it fully takes advantage of spectral, spatial, and phenology features of crops from satellite data. The classification experiments were conducted on crops of corn, soybeans, and others. To verify the effectiveness of the proposed method, comparisons with models of SVM, SegNet, CNN, LSTM, and different combinations were also conducted. The results show that our method achieved the best classification results, with the Kappa coefficient of 0.7933 and overall accuracy of 0.86. Experiments in other study areas also demonstrate the extensibility of the proposed method.

Keywords: adversarial generative network; crop classification; deep learning; multispectral remote sensing



Citation: Li, J.; Shen, Y.; Yang, C. An Adversarial Generative Network for Crop Classification from Remote Sensing Timeseries Images. *Remote Sens.* **2021**, *13*, 65. <https://dx.doi.org/10.3390/rs13010065>

Received: 29 October 2020

Accepted: 22 December 2020

Published: 26 December 2020

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

An in-depth understanding of crop types and its corresponding cultivated areas in national agricultural zones is necessary for agricultural monitoring and food security assessment [1], and contributes to the formulation and implementation of relevant policies. Modern earth observation (EO) programs can generate a large amount of remote sensing data, recording the characteristics of electromagnetic waves reflected by ground objects [2], and satellite remote sensing is widely used to obtain vegetation phenological characteristics [3]. Each feature has a unique spectral characteristic, which varies slightly with environmental factors, the accuracy of the sensing device, etc. The effective combination of different sensor sources [4,5], wavebands, and time-stamped remote sensing images provides more comprehensive feature information about crops. Therefore, it is a feasible and meaningful study to explore the classification of crops based on remote sensing images.

Traditional remote sensing-based image classification algorithms of machine learning are gradually being used in the classification [6] and recognition of remote sensing images. These algorithms can be divided into unsupervised and supervised categories [7]. The former contains algorithms such as K-means and ISODATA clustering. However, with the demand for agriculture classification based on complex terrain increasing gradually, those algorithms can no longer meet the accuracy of remote sensing image classification [2]. The latter contains maximum likelihood, minimum distance, and support vector machine (SVM). At this stage, SVM [8] has been widely used in remote sensing image classification,

while some problems also appeared. For example, SVM has a greater probability of leakage and misallocation when the sample size is large [9].

Deep learning, referring to deep neural networks, is a kind of machine learning algorithm and has been widely used due to its powerful feature extraction and data expression ability [10]. In recent years, the recognition rate of deep learning on most traditional recognition tasks has increased significantly [11]. A large number of studies have shown that deep learning can extract features from remote sensing images, and improve the classification accuracy [12]. The mainstream deep learning models, including deep belief network (DBN) [13], convolutional neural network (CNN) [14], stack autoencoders (SAE) [15], and recurrent neural network (RNN) [16], have been successfully used in applications of remote sensing-based classification in most cases. In particular, aiming at the phenomenon of “same matter and different spectrum” in remote sensing images, Wu and Yang [17] proposed recording spatial information combined with spectral information and used DBN to complete high-precision classification. Cao et al. [18] proposed an active deep learning method combining active learning and deep learning to train the CNN to classify on the hyperspectral image, which achieves better performance with fewer marker samples than traditional hyperspectral imagery (HSI) classification algorithms. Hsieh and Kiang [19] compared and validated the ability of CNN to complete crop classification on the HSI data in the Salinas Valley and mixed vegetation agriculture of Indian pine. Liang et al. [20] proposed a remote sensing image classification method based on a stacked noise reduction automatic encoder, which has a total classification accuracy of 95.7% on the remote sensing data of the Gaofen 1 (GF-1) satellite. From the previous studies, it can be found that CNN can handle the spatial autocorrelation problem well, but it cannot effectively deal with the long-term and complex time dependence [21]. Meanwhile, in the crop classification, phenology is very important in the accuracy improvement. RNN [2] is a neural network model specially used to deal with time correlation [22] and has been proven to be effective in many fields, such as speech recognition [23], natural speech processing [24], etc., especially the Long-Short Term Memory RNN (LSTM) [25]. Ndikumana et al. [26] used RNN to classify Sentinel-1 radar images for agricultural mapping, and the classification accuracy on rice reaches 96%. Sun et al. [27] built an end-to-end framework for crop mapping based on LSTM and achieved high accuracy, which promoted the classification of time series remote sensing images. However, LSTM is limited in dealing with spatially correlated data [28]. Goodfellow’s adversarial generation network (GAN) also shows amazing results when completing classification tasks [29]. Its training process combines inference and generative learning [30] to promote the model to better learn the potential probability distribution of given data. As far as we know, there are rare studies to combine the above three models to deal with the crop classification from multi-band and multi-temporal images.

In this study, we propose a model embedding LSTM and CNN into GAN to perform crop classification tasks. We conducted experiments on three classes (corn, soybeans, and other crops), and made comparisons with models of SVM, SegNet, CNN, LSTM, and different combinations to verify the effectiveness of the proposed method. Experiments in other study areas were also conducted to validate the extensibility. We expect that this method can provide technical support for agricultural monitoring and management.

2. Materials and Methods

2.1. Study Area and Datasets

We selected multiple study areas in the United States, as Figure 1 shows, these areas located in Iowa, Ohio, and Pennsylvania, with a total area of 437.88 km². The main crops in these areas are corn and soybeans. We categorized a few other crops and non-agricultural areas into a third category. When doing long-term exploratory experiments, taking into account the imaging time of Landsat8 and the crop growth period (US soybeans are sown in May–June, and harvested in October, and US corn is sown in March–April, and harvested in October), we choose the sequence time as July 22nd, August 7th, and September 24th.

The proportions of the three categories were 39.93%, 29.95%, and 30.12% in the three time-nodes, respectively.

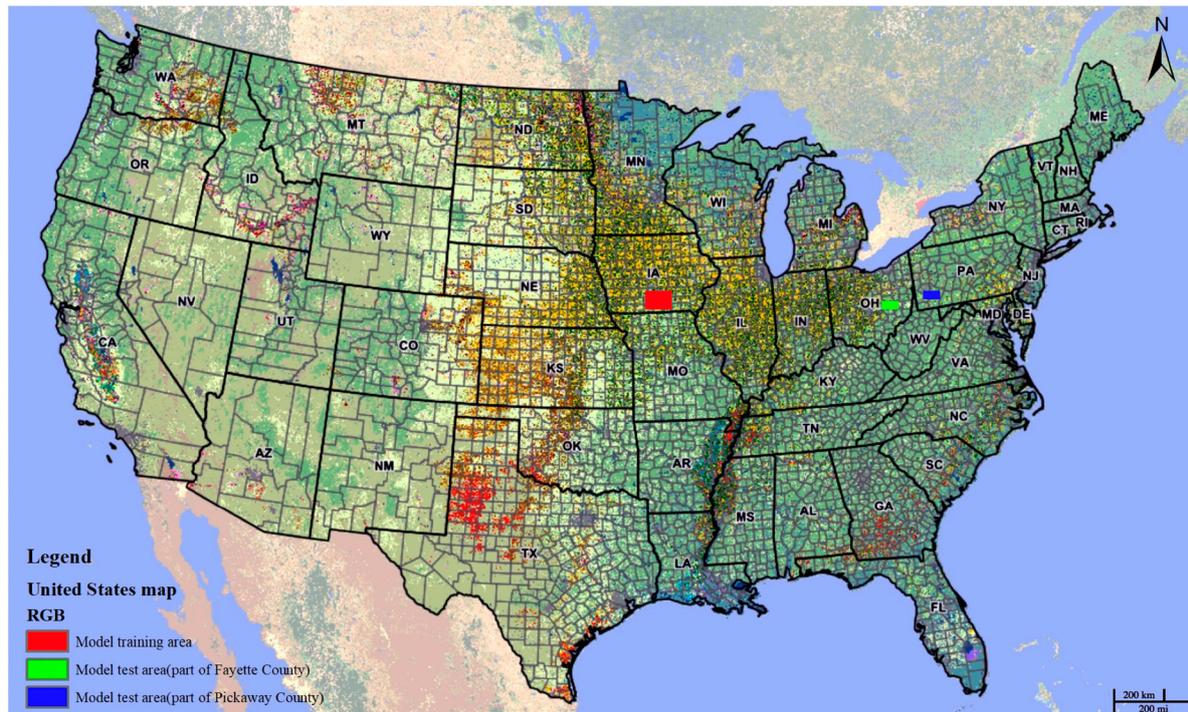


Figure 1. Study areas concentrated in Iowa, Ohio, and Pennsylvania States.

In these regions, the quantity of various categories is balanced, which is conducive to classification training. That is, when it comes to the classification task, the classification ability of the model can only be demonstrated on the premise that the sample size of each category is similar. Otherwise, it would be unfair for categories with small sample sizes, because the larger the sample size of a category, the more likely it is to influence the outcome of the final assessment.

In the experimental part, we used Tier 1 Landsat8 images in Iowa to train our model, and validated our model's generalization capability in Fayette County, Pennsylvania state and Pickaway County, Ohio state. The proportions of the three categories in the selected areas of Fayette County were 31.07%, 40.54%, and 28.39%, and the proportions of Pickaway County were 19.30%, 27.81%, 52.89% in three time-terms. All the image data were from the EarthExplorer website and processed by ArcGIS software with geometric, atmospheric, and radiation correction.

The label data were derived from Cropland Data Layer (CDL) data [31–33], high quality and robust crop type information at field scales from the U.S. Department of Agriculture (USDA) since 2008. The spatial resolution of CDL data is 30 m, the same as the Landsat8 images. In 2019, the Kappa coefficient of CDL data on corn and soybean was 0.889 and 0.835, respectively. Figure 2 shows part of the training data produced. Co-registration was done between the input and the label image.

We further split images into a series patches for the model training process. A fixed-size window (3×3) was used to cut the Landsat data produced to obtain a large number of remote sensing images with a size of 3×3 . The label for each remote sensing image was determined by the crop type represented by its center pixel at the corresponding location of the CDL image (corn is 0, soybeans are 1, and the rest is 2). Figure 3 visually shows the process of making the sample pair. We assign training and test sets at a 6:4 scale, resulting in a training set with 12,636 pairs of data and a test set containing 8424 pairs of data.

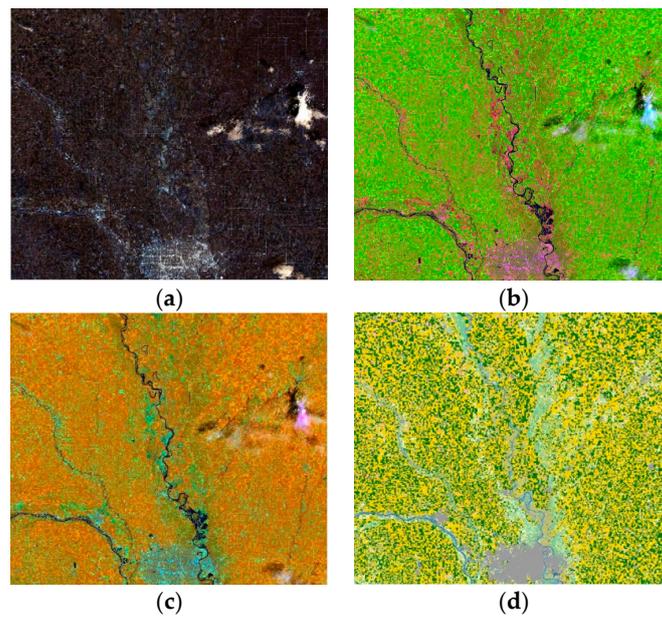


Figure 2. Part of the training data produced: (a) 432 band fusion image; (b) 652 band fusion map; (c) 564 band fusion map; (d) corresponding CDL data. (a–c) are obtained by Landsat 8 satellite on 7 August 2019.

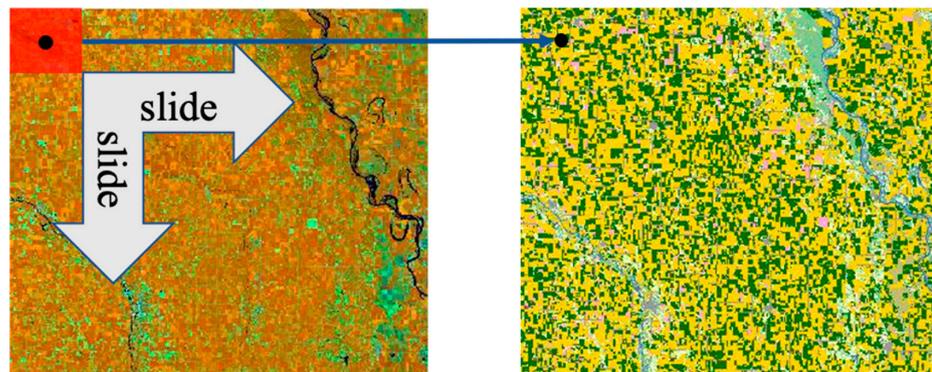


Figure 3. Example of making a training sample pair. The fixed-size window is slidingly cut on the input remote sensing image, and the crop type label corresponding to each remote sensing image is the crop with the center point at the corresponding position of the CDL image.

Figure 4 shows part of the training data produced from the combined data of bands 4, 3, and 2 on August 7. Our task was to classify the central pixels, and the remaining pixels in the image served as auxiliary information to help complete the classification task.

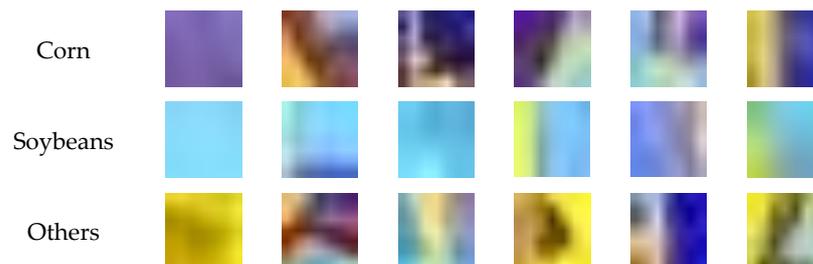


Figure 4. Part of the training examples.

2.2. GAN Embedded CNN and LSTM

Based on the GAN model, we combined LSTM and CNN to build the final classification model to process multi-temporal and multi-band fusion data. The model performs inference and generative training at the same time during training, and only uses the discriminator during testing. This section first reviews the basic principles and model architecture of LSTM and GAN, and finally gives our method.

The RNN takes sequence data as input, recursively in the direction of the sequence, and all nodes are connected in a chain [13]. That is, the output of the neuron at time $t-1$ will be the input of the neuron at time t together with the next input. Unlike traditional neural network models, RNN can take advantage of the time dependence of data. As a variant of RNN, LSTM solves the problem of gradient disappearance and gradient explosion when using RNN to build a deeper network. Equations (1)–(7) and Figure 5 show the internal computation of the LSTM unit. \odot stands for the Hadamard product, which multiplies the elements in two homotype matrices correspondingly. The input of LSTM is sequence variables $\{x_1, \dots, x_N\}$, where x_t is the feature vector and t is the corresponding timestamp. The γ^t in the formula represents the result of the splicing between the input at time t and the h^{t-1} passed from the previous state.

$$z = \tanh(\omega\gamma^t) \quad (1)$$

$$z^i = \text{sigmoid}(\omega^i\gamma^t) \quad (2)$$

$$z^f = \text{sigmoid}(\omega^f\gamma^t) \quad (3)$$

$$z^o = \text{sigmoid}(\omega^o\gamma^t) \quad (4)$$

$$c^t = z^f \odot c^{t-1} + z^i \odot z \quad (5)$$

$$h^t = z^o \odot \tanh(c^t) \quad (6)$$

$$Q^t = \text{sigmoid}(w'h^t) \quad (7)$$

The LSTM unit consists of two transfer states (memory state c^t and hidden state h^t) and three different gates (input gate z^i , forget gate z^f and output gate z^o). z^i determines how much input in the current state is retained, z^f determines how much information passed from the previous state needs to be forgotten, c^{t-1} . z^o determines how much information about the current state is passed to the next state. The three gated states not only determine the information flow inside the unit but also effectively prevent the problem of gradient disappearance and explosion. z as a temporary transfer state, and the hyperbolic tangent function is used to scale the input of the current state. h^t is calculated based on the newly obtained c^t , and γ^t acts on the newly obtained h^t to determine the output of the current state. The sigmoid and hyperbolic tangent functions used in the calculation process are calculated elementwise.

The adversarial generation network consists of two parts: generator G and discriminator D . The purpose of G is to learn the probability distribution of given data, and the input z is mapped to the data space $G(z; \theta_g)$ through a differentiable network. The essence of D is a classifier and the purpose is to correctly distinguish the data from G or the training set. G and D realize joint optimization in mutual games. Goodfellow et al. [29] proved that when the model reaches the global optimal value, the spatial distribution of G output data is equivalent to the real data distribution.

Our core idea to implement the method is to embed CNN and LSTM on the GAN model. Specifically, the discriminator is the classifier that finally completes the classification task, and CNN and LSTM are the feature extractors to extract the spatial and temporal features of the remote sensing image as the auxiliary input of the discriminator. Salimans et al. [34] once applied GAN to semi-supervised classification problems and achieved good results. Different from the general GAN, D not only needs to judge the

authenticity of the image but also needs to judge the category. If the classification task requires distinguishing category k , then the output layer of D will have $k + 1$ neurons. Roy et al. [35] argued that it is precisely because D needs to further extract the characteristics of the generated data when making true and false judgments, leading that GAN can achieve better results than Inception Net when completing the classification task.

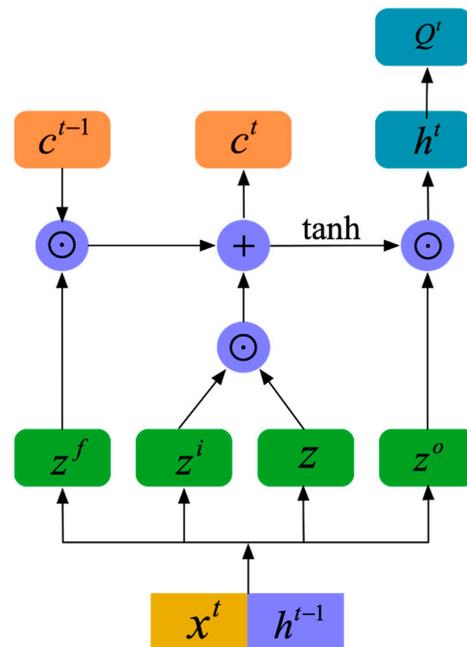


Figure 5. The internal structure of the long- and short-term memory neural network (LSTM) unit.

Salimans et al. [34] introduced a feature loss item and compares the features extracted from the real and fake images in the middle layer of the discriminator. This idea was extended by Johnson et al. [36] (perceptual loss term), which uses a certain layer of feature space of an external pre-trained network to construct the loss term. Inspired by this, we also used two pre-trained networks composed of CNN and LSTM to participate in the model construction. The difference is that the purpose of using external networks is to provide semantic information to help the discriminator judge rather than construct auxiliary loss items [37]. Specifically, we chose Inception Net [38] as the external network formed by CNN. When the input is a remote sensing image x , the extracted feature vector $s(x)$ is composed of the activation value extracted by the Inception Net. We built a simple network composed of three-layer LSTM units same as Lenco et al. [39] to extract time series features $t(x)$. Remote sensing images at different times constitute different states and are input into the network in time series. The input of the discriminator is a combination of x , $s(x)$, and $t(x)$.

The overall model architecture is shown in Figure 6. The LSTM part contains two cells used to extract the temporal characteristics of the images. The input data are divided into three sequences according to time and each sequence has a dimension of 49. We set the dimension of the output feature matrix as 3×3 , which is convenient to be spliced with the original image as the input of the generator. It should be noted that the Inception Net model corresponding to the CNN part was trained not only on the ImageNet dataset but also on 900 agricultural images (300 each for corn, soybean, and others) by simply modifying the input layer. During the training process, we also add the reshape and convolution operation for converting the feature matrix dimension of Mixed_7c layer to $2 \times 3 \times 3$ to concatenate with the original image. There is only one generator in our model and GAN chooses the similar generator architecture as MapGAN [40], which is mainly composed of the down-sampling layer, residual block, and up-sampling layer as shown in Figure 7. Each residual block contains two convolutional layers that do not change the dimensions

of the input data. We chose patchGAN as the discriminator, and the remote sensing image is judged in patches.

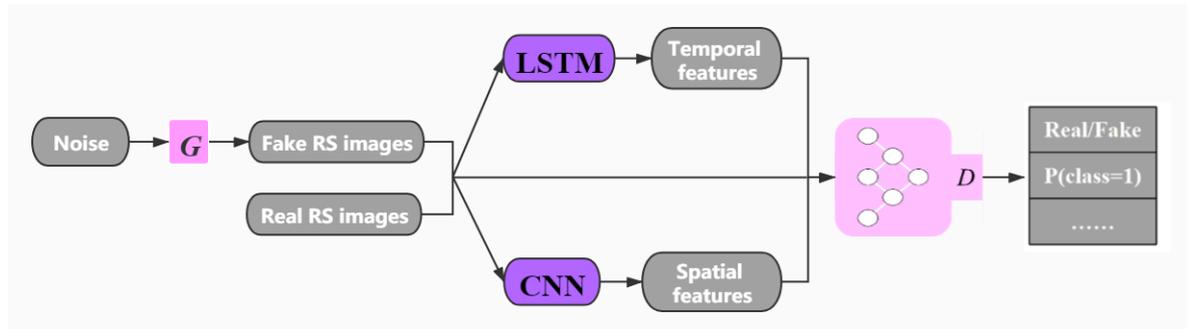


Figure 6. Overall model architecture.

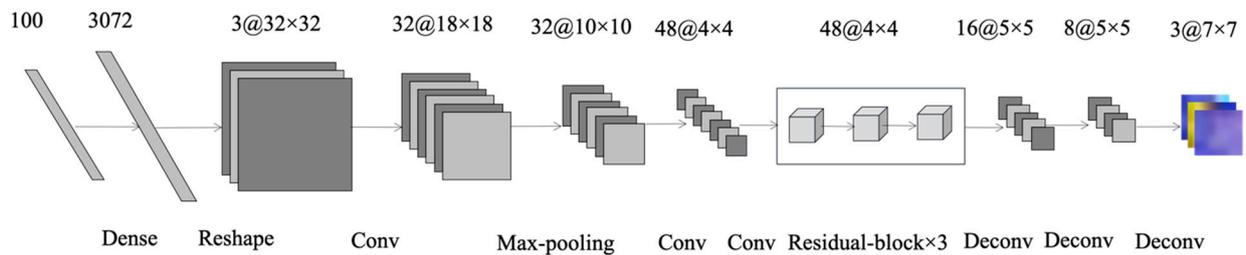


Figure 7. The generator architecture.

The output of the discriminator is a prediction of a multi-probability distribution represented by a $K + 1$ dimensional logit output which comprises of K real classes and the $(K + 1)$ class representing the fake images. We use y to represent the correct category of the discriminator corresponding to the image x . $P_D(y|D(x, s(x), t(x)), y < k + 1)$ means the probability that the image x belongs to the y category output by the discriminator. The least square loss function constructed by Mao et al. [41] is used to replace the original loss function of GAN to avoid the problem of gradient disappearance and further improve the performance of the model. The loss function of the generator is:

$$\min L(G) = \frac{1}{2} E_{z \sim p_z(z)} \left[(P_D(y|D(G(z), s(G(z)), t(G(z))), y < k + 1) - 1)^2 \right] \quad (8)$$

The loss function of the discriminator consists of a supervised loss item and an unsupervised loss item, namely $\min L(D) = L_{sup} + L_{unsup}$, where:

$$L_{sup} = -E_{x,y \sim p_{data}(x,y)} [\log(P_D(y|D(x, s(x), t(x)), y < k + 1))] \quad (9)$$

$$L_{unsup} = -\frac{1}{2} E_{z \sim p_z(z)} [\log(P_D(y = k + 1|D(G(z), s(G(z)), t(G(z)))))] - \frac{1}{2} E_{x,y \sim p_{data}(x,y)} [\log(1 - P_D(y = k + 1|D(x, s(x), t(x))))] \quad (10)$$

For the meaning of each symbol, please refer to the previous part of this paper.

3. Results and Discussion

In this study, we used Landsat8 satellite imagery and CDL data to conduct experiments to explore the performance of the model in three different categories (corn, soybeans, and others). Three experiments were conducted. The first experiment explored the best multi-band combination for crop classification when the input images are in a single time term. The second experiment applied the band combination method obtained in the first group to explore the classification capability of our model when the input images are in multi time-terms. We carried out a comparison among multiple sets of models (e.g., CNN, CNN

+ LSTM, LSTM, GAN + CNN, SVM, SegNet, GAN + CNN + LSTM). The third experiment tested our model to perform prediction classifications in the Fayette and Pickaway County and visualize the results.

3.1. Experiment Settings

We used the Inception Net model on behalf of the CNN model, and the three LSTM units are fully connected to the SoftMax layer to represent the LSTM as the baseline model. At the same time, we combined LSTM and CNN to get a new benchmark model, in which the combination method used the feature vector extracted from the pre-trained Inception Net network Mixed_7c layer as the auxiliary input of the LSTM unit. It is worth mentioning that we set the discriminator of GAN as CNN to implement the GAN+CNN model. The initial learning rate is set to 5×10^{-4} , and the linear decay rate is 5×10^{-5} . We used the RMSProp optimizer which the square root of the sum of historical gradient controlled by the attenuation coefficient, making the learning rate of each parameter different in the training process. The proportion of the three categories in the training set and test set is approximately 1.3:1:1. The model trained a total of 200 epochs, with a batch size of 32 and a training time of about 3 min per epoch. The experiment was carried out on the workstation with one GPU of NVIDIA M40, four CPUs of Inter Xeon Platinum 8163@2.5ghz, and RAM: 30GiB.

3.2. Band Combination Selection

To test which band combinations can provide the most abundant and useful information of crops in multi-temporal data classification, we investigated the degree of separation of three classes in different band combinations using ENVI software [42] before conducting classification experiments. The separability is represented by the Jeffries–Matusita distance. The larger the distance value is, the better the separability is, which is more beneficial for crop classification. Based on this idea, we explored the category separability of eight different band combinations at three different dates of 2019 (July 22, August 7, and September 24). It is worth noting that we explored the combination of the three bands to balance the relationship between speed and data volume. The experimental results are shown in Table 1.

Table 1. Jeffries–Matusita distance of the three categories in different band combinations at different times.

NO.	Band Combination	Crop Type	22 July 2019			7 August 2019			24 September 2019		
			Corn	Soybeans	Others	Corn	Soybeans	Others	Corn	Soybeans	Others
1	5, 6, 4	Corn	1.00	1.86	1.93	1.00	1.73	1.87	1.00	1.79	1.82
		Soybeans	1.86	1.00	1.76	1.73	1.00	1.74	1.79	1.00	1.85
		Others	1.93	1.76	1.00	1.87	1.74	1.00	1.82	1.85	1.00
2	5, 6, 2	Corn	1.00	1.61	1.79	1.00	1.77	1.83	1.00	1.62	1.84
		Soybeans	1.61	1.00	1.68	1.77	1.00	1.81	1.62	1.00	1.59
		Others	1.79	1.68	1.00	1.83	1.81	1.00	1.84	1.59	1.00
3	4, 3, 2	Corn	1.00	1.50	1.62	1.00	1.66	1.72	1.00	1.57	1.54
		Soybeans	1.50	1.00	1.57	1.66	1.00	1.68	1.57	1.00	1.71
		Others	1.62	1.57	1.00	1.72	1.68	1.00	1.54	1.71	1.00
4	5, 4, 3	Corn	1.00	1.76	1.68	1.00	1.67	1.83	1.00	1.55	1.62
		Soybeans	1.76	1.00	1.53	1.67	1.00	1.57	1.55	1.00	1.64
		Others	1.68	1.53	1.00	1.83	1.57	1.00	1.62	1.64	1.00
5	1, 2, 3	Corn	1.00	1.58	1.65	1.00	1.75	1.49	1.00	1.60	1.67
		Soybeans	1.58	1.00	1.47	1.75	1.00	1.56	1.60	1.00	1.53
		Others	1.65	1.47	1.00	1.49	1.56	1.00	1.67	1.53	1.00
6	5, 4, 1	Corn	1.00	1.21	1.83	1.00	1.42	1.56	1.00	1.51	1.68
		Soybeans	1.21	1.00	1.47	1.42	1.00	1.63	1.51	1.00	1.61
		Others	1.83	1.47	1.00	1.56	1.63	1.00	1.68	1.61	1.00
7	7, 4, 3	Corn	1.00	1.60	1.82	1.00	1.47	1.66	1.00	1.46	1.73
		Soybeans	1.60	1.00	1.39	1.47	1.00	1.28	1.46	1.00	1.51
		Others	1.82	1.39	1.00	1.66	1.28	1.00	1.73	1.51	1.00
8	7, 5, 4	Corn	1.00	1.25	1.56	1.00	1.59	1.57	1.00	1.37	1.64
		Soybeans	1.25	1.00	1.77	1.59	1.00	1.68	1.37	1.00	1.72
		Others	1.56	1.77	1.00	1.57	1.68	1.00	1.64	1.72	1.00

As can be seen from the table, the combination of 743, 753, and 541 has multiple Jeffries–Matusita distance values below 1.5 under three time-terms and has the worst comprehensive performance. The combination of 432 and 543 obtained a separation degree between 1.5 and 1.7, with moderate overall performance, and the 123 combination falls in between. The best-performing combinations are 564 and 562, with multiple Jeffries–Matusita distance values greater than 1.7. Among them, the separation degree of the 564 combination contains many values more than 1.8, which is the best in all the eight combinations. Interestingly, band 564 happens to be a classic combination for vegetation analysis. Therefore, the 564 band combination is more conducive to crop classification, and we will use this combination for subsequent experiments.

3.3. Accuracy Assessment of Crop Classification

To verify the ability of our method in dealing with the classification problem when the input data are time-series remote sensing images, we use Landsat8's 564 combination band images with three time-terms on July 22, August 7, and September 24 of 2019 to form series data. To maintain the consistency, the data generated by the GAN model are also in series in this experiment. In addition, we put together seven different models for comparison. The experimental results are shown in Table 2. The order of these models in performance effects is: LSTM < CNN < CNN + LSTM < GAN + CNN < SVM < SegNet < GAN + CNN + LSTM. The experiment testing the first four models can be regarded as ablation analysis. SVM and SegNet are two state of the art models at present. Our method (GAN+CNN+LSTM) achieves the highest Kappa coefficient (0.7933) and OA accuracy (0.86). It is worth noting that all models can achieve more than 85% accuracy in the classification of corn, but this is difficult for soybeans and other categories, which range mainly from 60% to 80%. The results of the first two models show that it is not enough to use only spatial features or temporal features when completing this classification task. The result of the model CNN+LSTM proves the effectiveness of combining the spatial and temporal features. Interestingly, the result of model GAN+CNN reflects the excellent ability of GAN in completing classification tasks. We consider that this is not only due to the fact that the discriminator itself is a CNN model, but the generation training process helps the model to use more false samples in the training process. The Kappa coefficient of both SVM and SegNet is above 0.7, showing the strong competitiveness. Compared to SegNet, our model improved by 3 percentage points.

Table 2. Classification accuracy of seven different models on multi-temporal and multi-band data sets.

NO.	Model	Crop Type	Confusion Matrix (%)			Kappa Coefficient	OA
			Corn	Soybeans	Others		
1	LSTM	Corn	92.36	0.63	7.01	0.3369	0.54
		Soybeans	62.67	32.53	4.80		
		Others	20.18	44.12	35.70		
2	CNN	Corn	89.54	2.51	7.94	0.4634	0.64
		Soybeans	26.21	68.19	5.60		
		Others	43.46	20.1	36.44		
3	CNN+LSTM	Corn	86.94	11.41	1.65	0.5140	0.68
		Soybeans	4.01	87.39	8.60		
		Others	4.68	78.04	17.28		
4	GAN+CNN	Corn	97.28	0.38	2.34	0.6498	0.77
		Soybeans	38.72	53.71	7.57		
		Others	12.23	7.63	80.14		
5	SVM	Corn	92.65	7.13	0.22	0.7222	0.81
		Soybeans	7.27	87.89	4.84		
		Others	4.65	30.41	64.94		
6	SegNet	Corn	96.21	1.46	2.33	0.7578	0.83
		Soybeans	12.59	70.02	8.39		
		Others	12.82	6.96	80.22		
7	GAN+CNN+LSTM	Corn	97.55	1.13	1.32	0.7933	0.86
		Soybeans	12.84	80.21	6.95		
		Others	10.33	11.24	78.43		

3.4. Mapping Classification Result

In order to visualize the model prediction effect more intuitively, we assigned the model prediction result to the corresponding color of the CDL image and then visualize it. Specifically, yellow represents corn, green represents soybeans, and black represents other categories. We used the seven models trained in Section 3.3 to make predictions and observations on the original images that include the training set and the test set. The results are shown in Figure 8, in which region 1 and region 2 are parts of region 3.

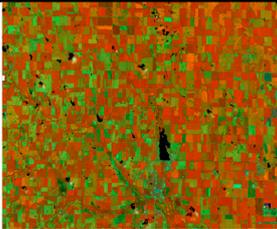
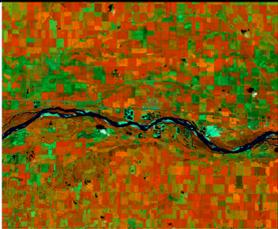
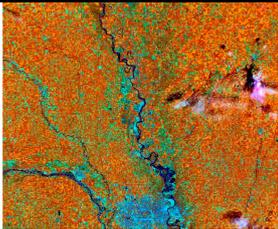
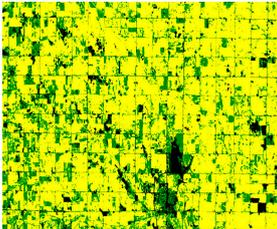
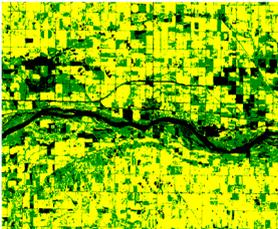
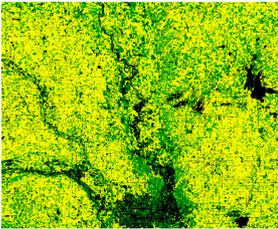
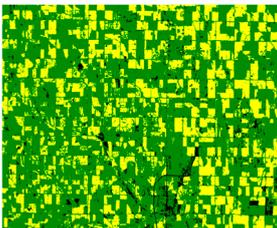
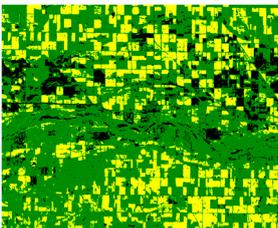
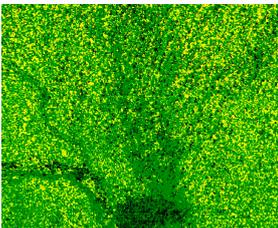
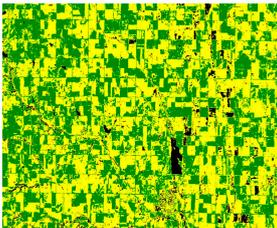
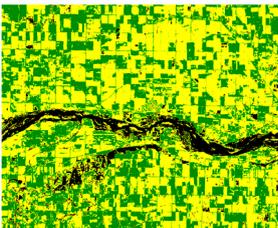
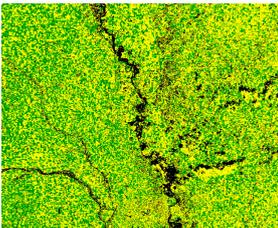
Data	Region1	Region2	Region3
RS image			
LSTM			
CNN			
LSTM+CNN			

Figure 8. Cont.

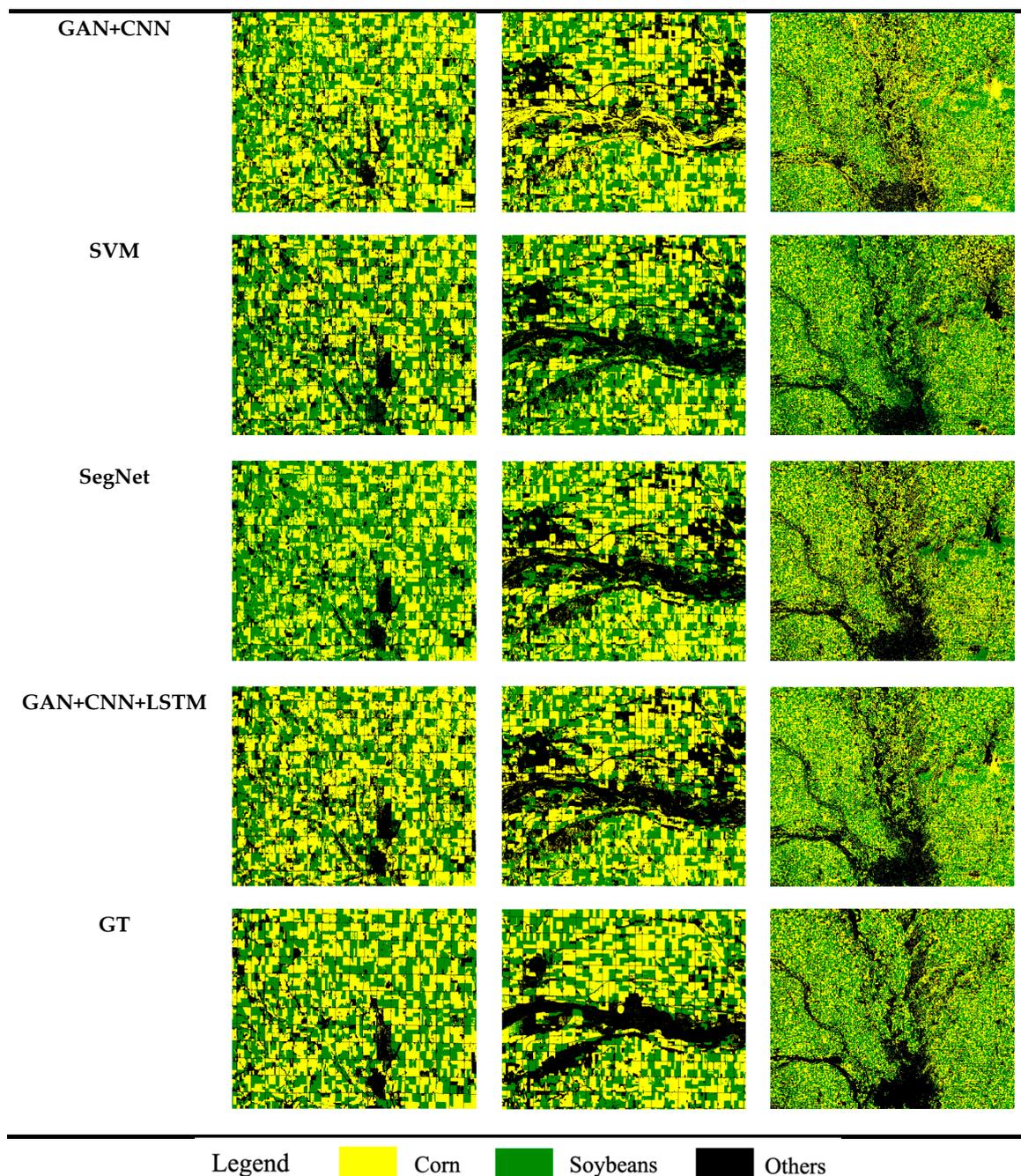


Figure 8. Visualization of the results of the model trained in the second experiment on the original image.

As we can see, the prediction results of LSTM are more confusing and the misclassification is serious, but the basic outline is shown. With the improvement of the performance of the model, the phenomena of misclassification and omission in a large area were significantly reduced, and the predict image gradually approached the ground truth image. It is worth noting that visually, our model has a visible improvement over the SVM and SegNet in the middle black area.

3.5. Model Extensibility Verification

In order to verify the generalization ability of the model, we selected new test areas from Fayette and Pickaway County. The classification category remains the same, and the time-series data are still the 564 band fusion images of Landsat8 satellite of the three months of July, August, and September. In addition, we used the SVM and SegNet model

for comparison considering their great performance in Section 3.4. The visualization results are shown in Figure 9. Compared with the experiment in Section 3.4, the difference between the models is more obvious in this section. The result obtained by our method has fewer noise points, more symmetrical color regions, and less misclassification. It can be seen that even for new regions, the model also has good classification capabilities. The quantitative evaluation results are shown in Tables 3 and 4. We can see that the three models performed better in Fayette County than Pickaway County, and the relative merits were the same as the results in Section 3.3. The OA accuracy obtained by our method is 0.85 and 0.88, respectively, which does not fluctuate greatly from the 0.86 obtained in Section 3.3.

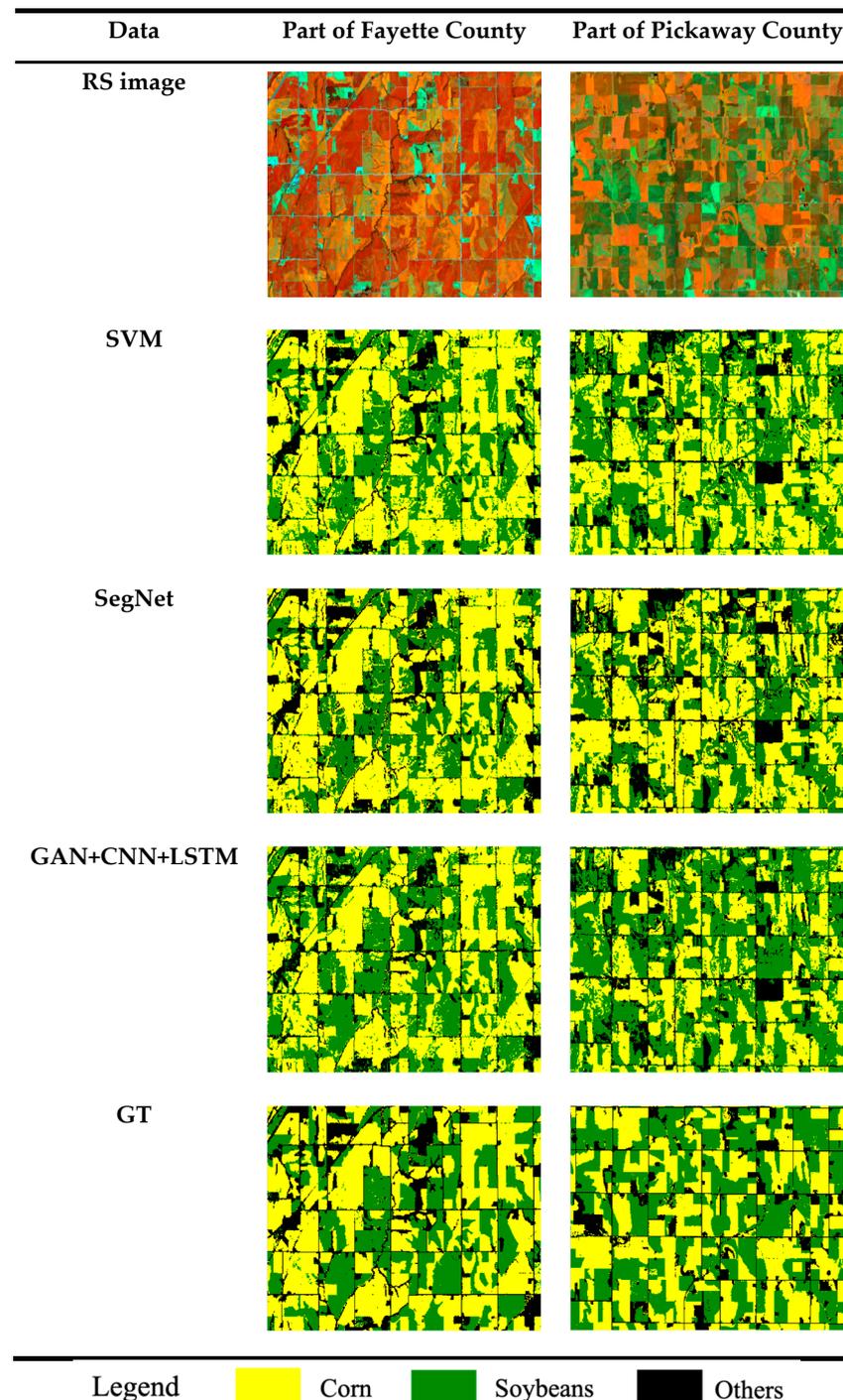


Figure 9. Visualization of the prediction results of the model in Fayette County and Pickaway County.

Table 3. Quantitative evaluation of the model's prediction results in Fayette County.

NO.	Model	Crop Type	Confusion Matrix (%)			Kappa Coefficient	OA
			Corn	Soybeans	Others		
1	SVM	Corn	88.53	7.14	4.33	0.6541	0.75
		Soybeans	17.3	71.83	10.87		
		Others	14.51	12.29	73.2		
2	SegNet	Corn	92.17	4.25	3.58	0.7186	0.81
		Soybeans	16.5	72.14	11.36		
		Others	10.39	19.27	70.34		
3	GAN+CNN+LSTM	Corn	90.31	6.72	2.97	0.7635	0.85
		Soybeans	5.37	81.42	13.21		
		Others	9.34	13.8	76.86		

Table 4. Quantitative evaluation of the model's prediction results in Pickaway County.

NO.	Model	Crop Type	Confusion Matrix (%)			Kappa Coefficient	OA
			Corn	Soybeans	Others		
1	SVM	Corn	91.47	5.12	3.41	0.7426	0.83
		Soybeans	13.87	74.66	11.47		
		Others	7.91	21.72	71.37		
2	SegNet	Corn	93.95	3.08	2.97	0.7651	0.87
		Soybeans	15.91	77.82	6.27		
		Others	11.90	10.89	77.21		
3	GAN+CNN+LSTM	Corn	90.93	7.62	1.45	0.7701	0.88
		Soybeans	3.60	87.21	9.19		
		Others	3.04	23.52	73.44		

4. Conclusions

In this study, we proposed a new method that combines GAN, CNN, and LSTM models to classify crops of corn and soybeans from Landsat8 time-series images. The method is feasible on the condition that the training samples are small, and it takes advantage of spectral, spatial, and phenology features from satellite data. Experiments including band combination selection, accuracy assessment of crop classification, and model extensibility verification showed that: (1) bands of 564 of Landsat 8 are effective for crop classification; (2) our model in the study area has the highest Kappa coefficient (0.7933) and OA (86%); (3) the trained model derived from training areas in Iowa can be effectively applied to Fayette and Pickaway County, and the OA is greater than 83%. The results demonstrate that the proposed method can realize crop classification well and provide decision support for agricultural management.

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

Funding: This research was funded by the Key R&D Program of Hubei Province, China under Grant No. 2020AAA004.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: [<https://pan.baidu.com/s/1rCKX3Q6o5OLu7OBH0XtEHA> with secret xtj2].

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. Zhai, Y.; Wang, N.; Zhang, L.; Hao, L.; Hao, C. Automatic crop classification in northeastern China by improved nonlinear dimensionality reduction for satellite image time series. *Remote Sens.* **2020**, *12*, 2726. [CrossRef]
2. Bin, W.A.N.G.; Doglin, F.A.N. Research progress of deep learning in classification and recognition of remote sensing images. *Bull. Surv. Mapp.* **2019**, *2*, 99–102.
3. Reed, B.C.; Brown, J.F.; VanderZee, D.; Loveland, T.R.; Merchant, J.W.; Ohlen, D.O. Measuring phenological variability from satellite imagery. *J. Veg. Sci.* **1994**, *5*, 703–714. [CrossRef]
4. Zhang, J. Multi-source remote sensing data fusion: Status and trends. *Int. J. Image Data Fusion* **2010**, *1*, 5–24. [CrossRef]
5. Wulder, M.A.; Loveland, T.R.; Roy, D.P.; Crawford, C.J.; Masek, J.G.; Woodcock, C.E.; Allen, R.G.; Anderson, M.C.; Belward, A.S.; Cohen, W.B.; et al. Current status of Landsat program, science, and applications. *Remote Sens. Environ.* **2019**, *225*, 127–147. [CrossRef]
6. Kotsiantis, S.B.; Zaharakis, I.; Pintelas, P. Supervised machine learning: A review of classification techniques. *Emerg. Artif. Intell. Appl. Comput. Eng.* **2007**, *160*, 3–24.
7. Mohd, H.I.; Pakhriazad, H.Z.; Shahrin, M.F. Evaluating supervised and unsupervised techniques for land cover mapping using remote sensing data. *Geogr. Malays. J. Soc. Space* **2009**, *5*, 1–10.
8. Bazi, Y.; Melgani, F. Toward an optimal SVM classification system for hyperspectral remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 3374–3385. [CrossRef]
9. Hui, C.; Huicheng, L. An improved BP neural network algorithm and its application. *Comput. Simul.* **2007**, *24*, 75–77.
10. Liu, B.; Du, S.; Du, S.; Zhang, X. Incorporating deep features into GEOBIA paradigm for remote sensing imagery classification: A patch-based approach. *Remote Sens.* **2020**, *12*, 3007. [CrossRef]
11. Gaoming, X.; Kai, H. The application of Tsinghua Sunway EPS software in the census of geographical conditions. *Surv. Spat. Geogr. Inf.* **2014**, *37*, 198–200.
12. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]
13. Chen, Y.; Zhao, X.; Jia, X. Spectral-spatial classification of hyperspectral data based on deep belief network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2381–2392. [CrossRef]
14. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Int. Neural Inf. Process. Syst.* **2012**, *60*, 1097–1105. [CrossRef]
15. Masci, J.; Meier, U.; Cireşan, D.; Schmidhuber, J. Stacked convolutional auto-encoders for hierarchical feature extraction. In *International Conference on Artificial Neural Networks, Proceedings of the 21st International Conference on Artificial Neural Networks, Espoo, Finland, 14–17 June 2011*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 52–59.
16. Zhang, D.; Wang, D. Relation classification via recurrent neural network. *arXiv* **2015**, arXiv:1508.01006.
17. Kaiyu, W.; Huihua, Y. Hyperspectral remote sensing image classification based on landmark spatial information. *Vedio Eng.* **2017**, *41*, 69–73.
18. Cao, X.; Yao, J.; Xu, Z.; Meng, D. Hyperspectral Image Classification With Convolutional Neural Network and Active Learning. *IEEE T. Geosci. Remote* **2020**, *58*, 4604–4616. [CrossRef]
19. Hsieh, T.-H.; Kiang, J.-F. Comparison of CNN algorithms on hyperspectral image classification in agricultural lands. *Sensors* **2020**, *20*, 1734. [CrossRef]
20. Liang, P.; Shi, W.; Zhang, X. Remote sensing image classification based on stacked denoising autoencoder. *Remote Sens.* **2018**, *10*, 16. [CrossRef]
21. Baatz, M.; Schäpe, A. Multiresolution Segmentation: An Optimization Approach for High Quality Multi-Scale Image Segmentation. Available online: http://www.agit.at/papers/2000/baatz_FP_12.Pdf (accessed on 25 December 2020).
22. Ebrahimi, J.; Dou, D. Chain based RNN for relation classification. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Denver, CO, USA, 31 May–5 June 2015; pp. 1244–1249.
23. Graves, A.; Mohamed, A.R.; Hinton, G. Speech recognition with deep recurrent neural networks. In Proceedings of the Advances in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–30 May 2013; pp. 6645–6649.
24. Linzen, T.; Dupoux, E.; Goldberg, Y. Assessing the ability of LSTMs to learn syntax-sensitive dependencies. *Trans. ACL* **2016**, *4*, 521–535. [CrossRef]
25. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A Search Space Odyssey. *arXiv* **2015**, arXiv:1502.04390. Available online: <https://arxiv.org/abs/1502.04390v1> (accessed on 12 February 2015). [CrossRef] [PubMed]
26. Ndikumana, E.; Ho Tong Minh, D.; 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]
27. Sun, Z.H.; Di, L.P.; Fang, H. Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series. *Int. J. Remote Sens.* **2019**, *40*, 593–614. [CrossRef]
28. Zhang, G.; Rui, X.; Poslad, S.; Song, X.; Fan, Y.; Ma, Z. Large-Scale, Fine-Grained, Spatial, and Temporal. *Sensors* **2019**, *19*, 2156. [CrossRef]

29. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
30. Zeng, J.; Wu, Y.; Liu, J.G.; Wang, L.; Hu, J. Learning and inference on generative adversarial quantum circuits. *Phys. Rev. A* **2019**, *99*, 052306. [[CrossRef](#)]
31. USDA National Agricultural Statistics Service Cropland Data Layer. Published Crop-Specific Data Layer. Verified USDA-NASS, Washington, DC. 2019. Available online: <https://nassgeodata.gmu.edu/CropScape/> (accessed on 10 May 2020).
32. Johnson, D.M. Using the Landsat archive to map crop cover history across the United States. *Remote Sens. Environ.* **2019**, *232*, 111286. [[CrossRef](#)]
33. Gao, F.; Anderson, M.C.; Zhang, X. Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sens. Environ.* **2017**, *188*, 9–25. [[CrossRef](#)]
34. Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved techniques for training gans. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Barcelona, Spain, 4–9 December 2016; pp. 2234–2242.
35. Roy, S.; Sangineto, E.; Sebe, N.; Demir, B. Semantic-fusion gans for semi-supervised satellite image classification. In Proceedings of the 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 684–688.
36. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 694–711.
37. Bazi, Y.; Al Rahhal, M.M.; Alhichri, H.; Alajlan, N. Simple yet effective fine-tuning of deep CNNs using an auxiliary classification loss for remote sensing scene classification. *Remote Sens.* **2019**, *11*, 2908. [[CrossRef](#)]
38. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2818–2826.
39. Ienco, D.; Gaetano, R.; Dupaquier, C.; Maurel, P. Land cover classification via multitemporal spatial data by deep recurrent neural networks. *IEEE Geosci Remote Sens. Lett.* **2017**, 1685–1689. [[CrossRef](#)]
40. Li, J.; Chen, Z.; Zhao, X.; Shao, L. MapGAN: An intelligent generation model for network tile maps. *Sensors* **2020**, *20*, 3119. [[CrossRef](#)]
41. Mao, X.D.; Li, Q.; Xie, H.R.; Raymond, Y.K.; Wang, Z.; Smalley, S.P. Least Squares Generative Adversarial Networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2794–2802.
42. Liu, X.K.; Zhai, H.; Shen, Y.L.; Lou, B.K.; Jiang, C.M.; Li, T.Q.; Hussain, S.; Shen, G.L. Large-scale crop mapping from multisource remote sensing images in google earth engine. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 414–427. [[CrossRef](#)]