OB-ConvLSTM: A sequential remote sensing crop classification model with OBIA and ConvLSTM models

Chen Luo
College of Information and Electrical Engineering
China Agricultural University
Beijing, China

Haiyang Li
College of Information and Electrical Engineering
China Agricultural University
Beijing, China

Jie Zhang
College of Information and Electrical Engineering
China Agricultural University
Beijing, China

Yaojun Wang*
College of Information and Electrical Engineering
China Agricultural University
Beijing, China
wangyaojun@cau.edu.cn

2023 11th International Conference on Agro-Geoinformatics (Agro-Geoinformatics) | 979-8-3503-0351-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/Agro-Geoinformatics59224.2023.10233296

***Abstract*—Remote sensing technology has become increasingly important in recent years due to its ability to collect high-resolution images of agricultural fields. One of the most popular methods for crop classification in agricultural fields is object-based image analysis (OBIA). At the same time, the convolutional long short-term memory (ConvLSTM) network has shown great potential in processing spatiotemporal data.**

In this study, we proposed a new model called OB-ConvLSTM (Object-based ConvLSTM) that combines OBIA and ConvLSTM for spatiotemporal crop classification tasks. This model extracts crop spectral information from the spatial dimension of remote sensing images, extracts crop growth information from multiple remote sensing images in the temporal dimension, and synthesizes the spatial and temporal dimension information to improve crop classification accuracy. Compared with traditional crop classification models based on single temporal remote sensing, the model proposed in this study is superior to existing models in classification accuracy and model robustness. The proposed OB-ConvLSTM model has been applied to crop classification tasks in major crop-producing regions, achieving over 93% of the crop species recognition accuracy, with mIoU reaching 83%.

The main contribution of this study is to design a temporal remote sensing image semantic segmentation model structure suitable for field crop classification, combining the OBIA method with ConvLSTM and improving the model's performance by optimizing model components such as activation functions and optimizers. Specifically, there are several innovations in the following aspects: First, to facilitate model input, this study uses the SLIC algorithm to segment remote sensing images into uniformly sized superpixel objects and aligns the superpixel objects in the temporal dimension; Subsequently, we used the ConvLSTM model totrain and classify superpixel objects with temporal information, and adopted the Mish activation function further to improve the semantic segmentation accuracy of remote sensing images. The dataset used in the experiment is temporal remote sensing images of corresponding ground crop types, including soybean and corn crops.

Our experimental results demonstrate that the proposed method achieved an overall classification accuracy of 93%, with a mean intersection-over-union (mIoU) of 83%. This model achieves higher accuracy than existing methods in field crop classification tasks. Future research can further expand the application of this model in crop yield prediction, crop management, and agricultural decision-making.

Keywords*—*sequential remote sensing, object-based image analysis, crop classification, long short-term memory

I. INTRODUCTION

Crop classification plays a vital role in monitoring crop growth, estimating yield, assessing disasters, and ensuring national food security [1-2]. Currently, the classification of crops using satellite remote sensing employs two primary approaches. The first strategy utilizes the spectral signature of a single satellite scene on a singular day within the growing season for classification [3-5]. Diverse land covers possess different reflectance spectral features, and machine learning or neural networks can extract these features and classify them. However, during the peak growing season when satellite images are usually obtained, the spectral information of certain crops is comparable, resulting in minimal spectral variation between crops and vegetation during specific periods of the year, thereby amplifying the complexity of crop classification. The second strategy comprehensively employs spectral and temporal information from one or more growing seasons [6-8]. This plan employs time series information of crops combined with spectral information for classification and uses deep learning models, such as recurrent neural network (RNN), long short-term memory network (LSTM), and other models for time series classification. These models extract features from remote sensing time series information, thereby enhancing the accuracy of crop classification.

Furthermore, the exponential growth of image data has been observed with the advent of high-resolution remote sensing and unmanned aerial vehicle (UAV)-based photogrammetry technology. Accommodating the processing of such data necessitates higher computing resources with enhanced arithmetic power and efficient algorithms. The conventional pixel-based imagery processing techniques are inadequate in enhancing feature recognition and classification accuracy. Moreover, the vast amount of feature information is highly detailed and the spectral features of different features possess a certain degree of similarity. This similarity results in a reduction of statistical separability of the image spectral domain. To address these challenges, numerous researchers have

979-8-3503-0351-3/23/$31.00 ©2023 IEEE

embraced the Object-Based Image Analysis [9](OBIA) research paradigm, which involves upgrading the analysis unit from pixels to objects and expanding the analysis granularity. This approach enables better utilization and extraction of texture information, spatial structure, and contextual relationships of remote sensing images. By utilizing the OBIA approach, researchers can improve the accuracy of feature classification, recognition, and distribution pattern analysis.

In this study, a novel OBIA modeling method called OB-ConvLSTM is proposed, and the ConvLSTM model [10] is combined with the OBIA method and applied to handle the feature classification problem of time-series remote sensing images, and experimental analysis is performed on the Landsat ARD dataset.

II. MATERIALS AND METHODS

A. Dataset

This study uses a dataset constructed based on Analysis Ready Data (ARD) collected by the US satellite Landsat. The dataset was downloaded from the official website of the United States Geological Survey (USGS) (https://www.usgs.gov/) and was always processed to the highest scientific standards and processing levels suitable for monitoring and assessing landscape change. The Crop Data Layer (CDL) is based on Landsat Thematic Mapper data, Common Land Unit data, NASS June Agriculture Survey data, and National Land Cover Dataset data after analysis and mapping, and its geospatial data products The CDL data are hosted on the CropScape website (https://nassgeodata.gmu.edu/CropScape/) server [11]. The CDL provides more than 100 land cover and crop categories, which were investigated by manual fieldwork and labeled with the corresponding categories. images with a spatial resolution of 30 m, which is consistent with the resolution of ARD data, and the accuracy of crop categories are high, especially for two major crops, maize and soybean, which have a classification accuracy of over 95%. Therefore, CDL data was selected as the labeled data for classification in this study.

The specific data are mainly from six regions of the U.S. corn and soybean planting belt. Two of these regions were selected for this study: their ARD grid coordinate system row numbers are h18v07, h17v07, and their approximate locations are shown in Figure 1.

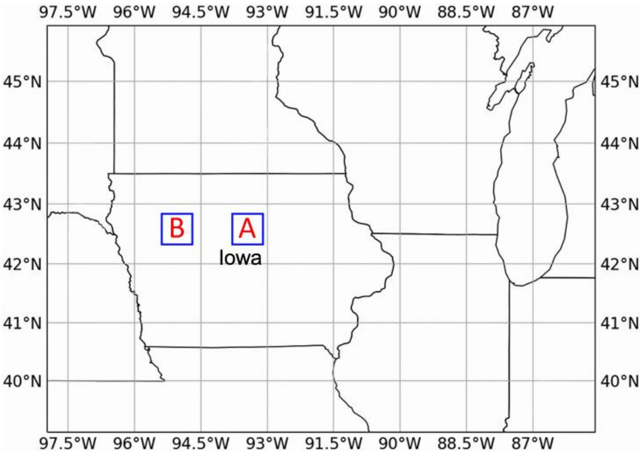

Fig. 1. Study area

In the U.S. Corn Belt, the growing cycles of corn and soybeans are from late April to mid-September and mid-May to late September, respectively. These areas typically have a single-season cropping system. Therefore, we used corn, soybean, and "other" as the three categories of interest and assigned corresponding crop category labels to each image of the crop spatial distribution map for each year.

B. Data Preprocessing

Surface reflection data from Landsat Analysis Ready Data (ARD) were downloaded from the USGS. The original remote sensing image has a size of 5000 × 5000 and a resolution of 30 meters. We obtained a 500×500 size image by cropping the original image to the center. Based on the U.S. state-level corn and soybean growth progress reports [12], the study assumed April 15 as the earliest planting date. Therefore, observation dates ranged from week 1 (April 22) to week 23 (September 23) after April 15, covering the growth of corn and soybeans from planting to maturity in these states. Therefore, April 22 to September 23 of each year from 2018 to 2021 were selected as observation dates in this study.

For the training of the classification model, six optical bands were selected as input variables from the surface reflection data for each observation date, including Blue, Green, Red, Near-Infrared, Shortwave Infrared 1, and Shortwave Infrared 2. However, the presence of missing data and invalid observations leads to inconsistent observation frequencies. To solve this problem, we used linear temporal interpolation to fill in the missing values, which is effective in previous studies [13]. Linear interpolation is performed based on the most recent valid observations before and after the target time point of each image element to obtain a time series consisting of multiple equally spaced remote sensing observations. The whole data preprocessing process is shown in Figure 2.

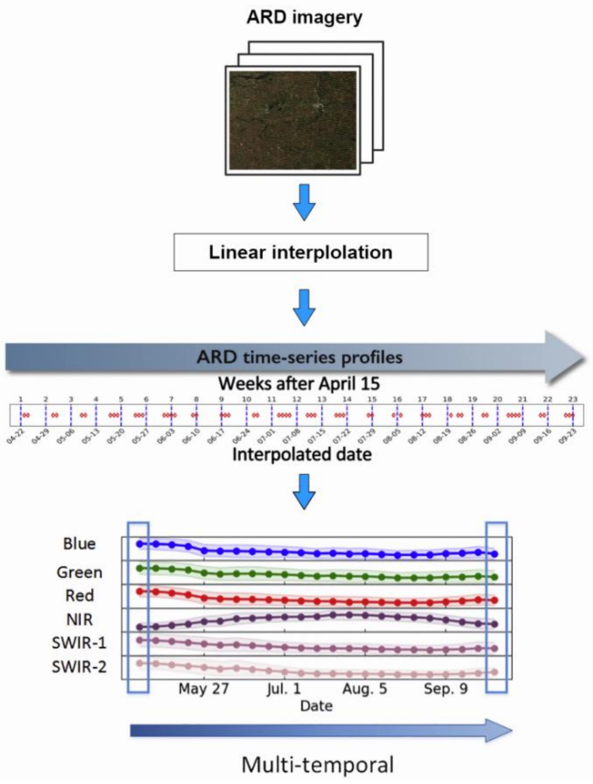

Fig. 2. Data pre-processing

C. *Overview of OBIA Analysis Methods*

OBIA is an object-level geoscientific analysis method, which is currently widely used in the field of remote sensing for feature extraction and object classification at the geographic object level. The basic idea is to divide the remote sensing image into several non-overlapping and unique regions through image segmentation, and then use these segmented regions as the basic unit of analysis for subsequent processing. It can enhance the separability between objects with the same object with different spectrums and different types with the same spectrum, avoid the uncertainty caused by mixed pixels in the fuzzy boundary, and improve the anti-interference ability to noise. At the same time, OBIA can significantly reduce the number of analysis primitives, thereby greatly improving analysis efficiency. Therefore, in the classification task of high-resolution remote sensing images, OBIA has high precision and practicability and is an important technological process of remote sensing image analysis.

D. *SLIC Superpixel Segmentation Algorithm*

Simple Linear Iterative Cluster [14] (SLIC) is an efficient superpixel segmentation method that can segment images into homogeneous and compact superpixel objects. Traditional segmentation methods such as mean drift and multiscale segmentation generate segmented objects with very large differences in size and shape. If the segmentation objects generated by these segmentation methods are directly input to CNN, different segmentation objects occupy different proportions of the input image, which leads CNN to mine too many features from the background area of the segmented objects and ignore some important information, which will eventually affect the training and classification results. To solve this problem, the superpixel segmentation algorithm SLIC is used to efficiently segment images into uniform and compact superpixel objects, thus improving computational efficiency while maintaining accuracy.

In this experiment, we use the Skimage library to implement the SLICO algorithm and apply it to the segmentation task of remote sensing images. By introducing the adaptive compactness measure of SLICO, remote sensing images can be segmented more efficiently and accurately, so that useful information can be extracted and utilized.

E. *ConvLSTM*

ConvLSTM is a deep learning model that combines a convolutional neural network and a long short-term memory network (LSTM) to efficiently process both time-series data and spatial image data. The model is not only capable of predicting future states by remembering previous information but also captures the spatiotemporal relationships in the input data. Therefore, ConvLSTM has shown good performance in many sequence prediction tasks, such as video analysis and weather prediction.

Compared with traditional LSTM, ConvLSTM can better capture the spatiotemporal features in images and videos, thus improving the prediction effect.

The core idea of ConvLSTM is to convert the input data into three gate units (input gate, forget gate, output gate) at each moment node of the LSTM network, and update the cell state (cell state). This design enables ConvLSTM to effectively model the long-term dependencies of the input data and simultaneously consider the correlation between different locations and times of the input data. In addition, ConvLSTM also introduces convolution operations, which are usually applied to each gating element and state output, ensuring that the model can capture the spatial and temporal characteristics of the input data. ConvLSTM also supports multi-layer stacking, and each layer can set different convolution kernel sizes and strides.

F. *OB-ConvLSTM Modeling Process*

First, for the pre-processed temporal remote sensing images, the SLICO algorithm is used to segment them by superpixels to obtain a series of superpixel objects with similar shapes and close areas. Since the original image has temporal dimensional information, each superpixel object also retains the temporal information. To satisfy the input requirements of the ConvLSTM model, each superpixel object needs to be cropped into a square region. Through several attempts, we choose to crop a square region of size 10x10 centered on the center of gravity of each superpixel object. The whole process of extracting the superpixel objects is shown in Figure 3.

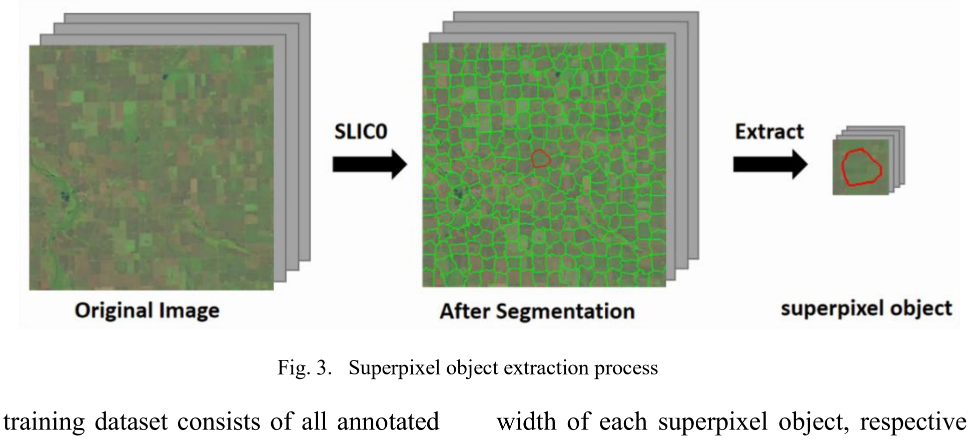

Fig. 3. Superpixel object extraction process

The superpixel training dataset consists of all annotated superpixel objects. Since each superpixel object contains information in the time dimension, this dataset can be represented as a 5D tensor: [B, S, F, H, W]. Among them, B represents the batch size, S represents the length of the time series (23 in this paper), F represents the number of features (6 in this paper), and H and W represent the height and width of each superpixel object, respectively. To train and classify superpixel objects, we feed this 5D tensor into a ConvLSTM model. After the model is trained, the category of each superpixel object can be obtained. Then, apply the class label value of each superpixel object to every pixel within that superpixel. All superpixel objects are concatenated according to their positions in the original image, and the predicted label map can be obtained. Finally, we optimize the predicted label map using a conditional random field [15] (CRF) to get the final output. The modeling process of the entire OB-ConLSTM is shown in Figure 4.

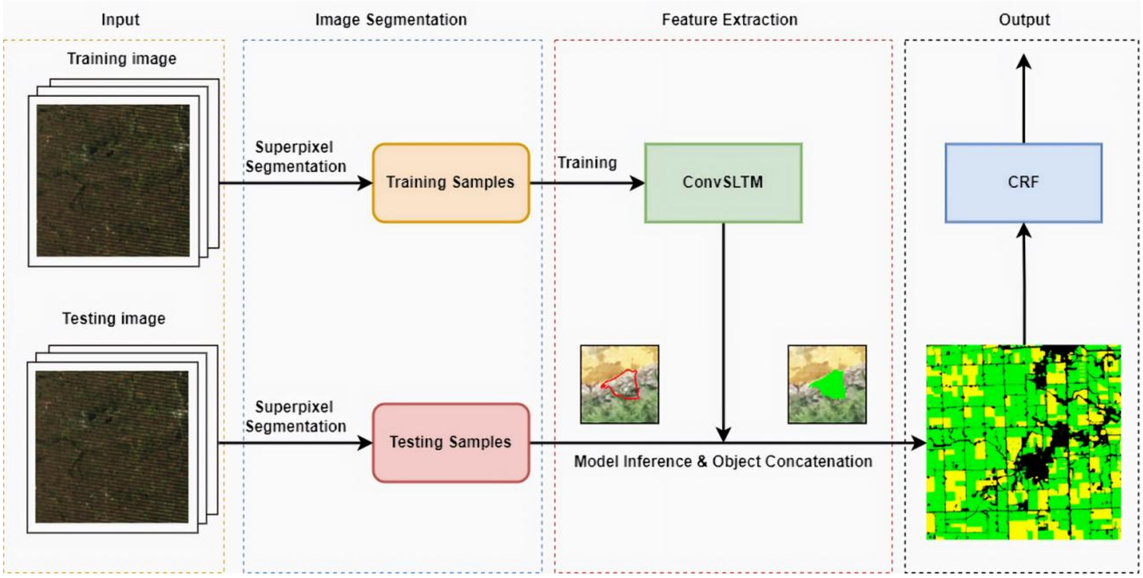

Fig. 4. Modeling process of OB-ConvLSTM

The method is established to realize the theoretical and technical integration of OBIA and ConvLSTM. By using the superpixel segmentation object as a bond, the use of ConvLSTM network is able to extract the temporal and spatial features of the segmented object, thus improving the final classification accuracy. In addition, the application of superpixel segmentation effectively avoids the pretzel phenomenon caused by mixed image elements in fuzzy boundaries and can significantly reduce the computational effort and the required storage space of the algorithm.

III. RESULTS AND DISCUSSION

A. *Experimental Environment and Experimental Parameters*

We use SGD as the optimizer to optimize the model parameters and set its momentum parameter to 0.9. At the same time, in order to better control the convergence speed and stability of the model, we use an adaptive learning rate regulator. Furthermore, we manually adjusted the learning rate (LR), mini-batch size, and training period (epoch) according to the evaluation results. Evaluation metrics include precision, recall, F1, and mIoU.

B. *Comparison of Object-based ConvLSTM and Pixel-based LSTM*

In the experiment, we selected two different sites (Site A and Site B) in the Iowa region as the research area and used three-year time-series remote sensing data from 2019 to 2020 as the training set. To improve the stability and reliability of the model, the training set is divided into five equal parts, and 1/5 of them is used as the verification set, while the test set uses time-series remote sensing data in 2021. Such an experimental design helps to make the model better adapt to remote sensing image data in different regions and periods, further improves the performance of the model, and ensures that it has good generalization ability in different regions and times.

In our experiments, we performed a comparative evaluation of two different temporal classification models. These two models are the OB-ConvLSTM model based on object level and the ordinary LSTM model based on the pixel level. We evaluate them using precision, recall, f1, mIoU, OA, and other indicators and present the results in Table 1 and Table 2.

TABLE I. COMPARISON OF CLASSIFICATION PERFORMANCE OF CONVLSTM AND LSTM IN SITE A REGION

Model	**Precison(%)**	**Recall(%)**	**F1(%)**	**MIoU(%)**	**OA(%)**	**Reasoning Speed（s）**
ConvLSTM	**92.86**	**92.99**	**92.85**	83.52	**92.99**	**5**
LSTM	92.32	92.02	92.04	**83.69**	92.02	20

TABLE II. COMPARISON OF CLASSIFICATION PERFORMANCE OF CONVLSTM AND LSTM IN SITE B REGION

Model	**Precison(%)**	**Recall(%)**	**F1(%)**	**MIoU(%)**	**OA(%)**	**Reasoning Speed（s）**
ConvLSTM	**93.30**	**93.45**	**93.27**	**83.92**	**93.45**	**5**
LSTM	93.14	92.50	92.66	83.86	92.50	20

The outcomes presented in Table 1 and Table 2 indicate that the OB-ConvLSTM model outperforms the LSTM model in both regions in terms of precision, recall, f1, mIoU, and OA. Notably, the performance gap is more substantial in the Site B area. This finding demonstrates the superiority of the OB-ConvLSTM technique that integrates object-based image analysis (OBIA) and convolutional long short-term memory (ConvLSTM) in accomplishing remote sensing image classification tasks. This method can more accurately extract feature information at the object level, resulting in higher classification accuracy.

Furthermore, the inference speed of the OB-ConvLSTM model is significantly faster than that of the LSTM model. This may be because OB-ConvLSTM performs feature extraction based on the object level, which reduces redundancy and makes the algorithm more efficient. Therefore, the OB-ConvLSTM method not only has higher classification accuracy but also has faster reasoning speed and wider application value.

In summary, the experimental results show that the OB-ConvLSTM method combining OBIA with ConvLSTM exhibits superior performance in remote sensing image classification tasks. It is especially advantageous in processing object-level time-series remote sensing images. Therefore, the OB-ConvLSTM method is an effective method worthy of popularization and application in the domain of remote sensing image classification.

C. Migration Capability Test of the Model

In this study, we fully explored the migration capability of the OB-ConvLSTM model. The model was first trained in region A using time-series remote sensing data from 2018 to 2020, and the training set was divided into five equal parts, of which 1/5 was used as the validation set. Subsequently, the trained model was applied to the test set in region B for classification testing, while the values of evaluation metrics (precision, recall, f1-score, mIoU, and OA) were recorded. Correspondingly, the models were also trained in region B using time-series remote sensing data from 2018 to 2020, and the training set was also divided into five equal parts, with 1/5 of them serving as the validation set. Then, the trained model was applied to the test set in region A for classification testing, and the values of relevant evaluation indexes were recorded. By comparing these two sets of results, we obtained the performance of the OB-ConvLSTM model for migration between different regions and evaluated the strength of its migration ability. The experimental results are shown in Table 3.

TABLE III. MIGRATION CAPABILITY TEST RESULTS OF OB-CONVLSTM

Base Site	Precison(%)	Recall(%)	F1(%)	MIoU(%)	OA(%)
A	91.98	92.02	91.91	82.16	92.02
B	92.24	92.39	92.27	82.23	92.39

Among them, Base Site indicates the region used for training. For example, when the Base Site is A, it means that the model is trained at Site A and tested at Site B.

Through the comparative analysis of the above experiments, we conclude that the OB-ConvLSTM model has a strong migration ability. When the model was trained at Site A and then migrated to Site B for testing, the accuracy and OA performed well, and the performance was slightly higher than that of the model trained at Site B. Similarly, when the model is migrated to Site A for testing after being trained at Site B, good classification results are also achieved. This shows that the method has a certain generalization ability between different regions, and can extend the classification accuracy of remote sensing images in a certain region to other regions, which has better practicability and generality.

From the experimental results, we can see that the classification performance of OB-ConvLSTM in Site A and Site B is basically equal with little difference, and its accuracy, mIoU, and OA indexes all reach a high level. Therefore, we can consider that the OB-ConvLSTM method is not only suitable for remote sensing image classification in a specific region but also has strong versatility and transferability.

To sum up, the OB-ConvLSTM method not only exhibits good remote sensing image classification performance in specific regions but also has strong migration ability, which can extend the remote sensing image classification accuracy in a certain region to other regions. This method provides reliable technical support and guidance for remote sensing image processing and analysis and has broad application prospects and development space.

IV. CONCLUTION

This study has conducted in-depth investigations on two issues: the first issue is that the single-temporal remote sensing crop classification is difficult to distinguish the spectral information of different crops in the same growing season; the second issue is that pixel-based remote sensing image segmentation is difficult to capture spatial contextual information. To address these issues, we tried a temporal remote sensing crop classification method based on the object level. The specific method is to combine the ConvLSTM model and OBIA technology and use the ARD dataset as input and the CDL dataset as the label to conduct experiments. First, the time series remote sensing images are generated by linear interpolation, and then the time series remote sensing images are segmented using the SLIC algorithm, and then the superpixel objects are cut into squares of the same size and formed into a batch input ConvLSTM model for training and classification. Finally, the segmentation results are further optimized using conditional random fields to obtain the final segmentation results.

The experimental results show that the OB-ConvLSTM method has better performance and migration ability in the object-level time-series remote sensing crop classification task. The method achieved an overall accuracy of over 92 percent across two different test regions. In addition, the OB-ConvLSTM method has a faster inference speed and a high application value for processing large-scale remote sensing images.

In summary, we believe that the method can not only improve the accuracy and efficiency of remote sensing crop classification but also can be applied to other remote sensing image classification and segmentation tasks.

Contribution

Chen Luo: Methodology, data analysis and writing; Haiyang Li:Formal analysis and writing; Jie Zhang: Review and editing; Yaojun Wang: Methodology, funding acquisition and writing.

References

[1] F. Qixin, Y. Liao, W. Weisheng, C. Tao, and H. Shuangyan, “CNN remote sensing crop classification based on time series spectral reconstruction,” Journal of University of Chinese Academy of Sciences, vol. 37, pp. 619-628, September 2020.

[2] S. Fatih and F. KOYUNCU, “Multi criteria decision analysis to determine the suitability of agricultural crops for land consolidation areas,” International Journal of Engineering and Geosciences, vol. 6, pp. 64-73, January 2021.

[3] T. G. Van Niel and T. R. McVicar, “Determining temporal windows for crop discrimination with remote sensing: a case study in south-eastern Australia,” Computers and electronics in agriculture, vol. 45, pp. 91-108, August 2004.

[4] C. Boryan, Z. Yang, R. Mueller, and M. Craig, “Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program,” Geocarto International, vol. 26, pp. 341-358, April 2011.

[5] C. Yang, J. H. Everitt, and D. Murden, “Evaluating high resolution SPOT 5 satellite imagery for crop identification,” Computers and Electronics in Agriculture, vol. 75, pp. 347-354, February 2011.

[6] J. Chang, M. C. Hansen, K. Pittman, M. Carroll, and C. DiMiceli, “Corn and soybean mapping in the United States using MODIS time-series data sets,” Agronomy Journal, vol. 99, pp. 1654-1664, November 2007.

[7] S. Foerster, K. Kaden, M. Foerster, and S. Itzerott, “Crop type mapping using spectral–temporal profiles and phenological information,” Computers and Electronics in Agriculture, vol. 89, pp. 30-40, November 2012.

[8] B. D. Wardlow, S. L. Egbert, and J. H. Kastens, “Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains,” Remote sensing of environment, vol. 108, pp. 290-310, June 2007.

[9] M. D. Hossain, and D. Chen, “Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective,” ISPRS Journal of Photogrammetry and Remote Sensing, vol. 150, pp. 115-134, April 2019.

[10] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, “Convolutional LSTM network: A machine learning approach for precipitation nowcasting,” Advances in neural information processing systems, vol. 1, pp. 802–810, 2015.

[11] C.Boryan, Z.Yang, R.Mueller, and M.Craig, “Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program,” Geocarto International, vol. 26, pp. 341–358, April 2011.

[12] Quick Stats 2.0, USDA-NASS, Washington, DC, 2018.

[13] J. Inglada, M. Arias, B. Tardy, H. Olivier, V. Silvia, D. Morin, et al, “Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery,” Remote Sensing, vol. 7, pp. 12356-12379, September 2015.

[14] A. Radhakrishna, S. Appu, S. Kevin, L. Aurelien, F. Pascal, and S. Sabine, “SLIC Superpixels,” EPFL, pp. 1-15, June 2010.

[15] L. John, M. Andrew, and C. N. P. Fernando, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” Proceedings of the 18th International Conference on Machine Learning 2001 (ICML 2001), pp. 282-289, June 2001.