

Article

Integrating the Continuous Wavelet Transform and a Convolutional Neural Network to Identify Vineyard Using Time Series Satellite Images

Longcai Zhao ^{1,2} , Qiangzi Li ^{1,*}, Yuan Zhang ¹, Hongyan Wang ¹  and Xin Du ^{1,3} 

¹ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; zhaolc@radi.ac.cn (L.Z.); zhangyuan@radi.ac.cn (Y.Z.); wanghy@radi.ac.cn (H.W.); duxin@radi.ac.cn (X.D.)

² University of Chinese Academy of Sciences, Beijing 100049, China

³ Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions (Ministry of Education), Henan University, Kaifeng 475004, China

* Correspondence: liqz@radi.ac.cn; Tel.: +86-010-6485-5094

Received: 20 August 2019; Accepted: 8 November 2019; Published: 12 November 2019



Abstract: Grape is an economic crop of great importance and is widely cultivated in China. With the development of remote sensing, abundant data sources strongly guarantee that researchers can identify crop types and map their spatial distributions. However, to date, only a few studies have been conducted to identify vineyards using satellite image data. In this study, a vineyard is identified using satellite images, and a new approach is proposed that integrates the continuous wavelet transform (CWT) and a convolutional neural network (CNN). Specifically, the original time series of the normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), and green chlorophyll vegetation index (GCVI) are reconstructed by applying an iterated Savitzky-Golay (S-G) method to form a daily time series for a full year; then, the CWT is applied to three reconstructed time series to generate corresponding scalograms; and finally, CNN technology is used to identify vineyards based on the stacked scalograms. In addition to our approach, a traditional and common approach that uses a random forest (RF) to identify crop types based on multi-temporal images is selected as the control group. The experimental results demonstrated the following: (i) the proposed approach was comprehensively superior to the RF approach; it improved the overall accuracy by 9.87% (up to 89.66%); (ii) the CWT had a stable and effective influence on the reconstructed time series, and the scalograms fully represented the unique time-related frequency pattern of each of the planting conditions; and (iii) the convolution and max pooling processing of the CNN captured the unique and subtle distribution patterns of the scalograms to distinguish vineyards from other crops. Additionally, the proposed approach is considered as able to be applied to other practical scenarios, such as using time series data to identify crop types, map landcover/land use, and is recommended to be tested in future practical applications.

Keywords: vineyard; identification; CWT; CNN; Sentinel-2; remote sensing

1. Introduction

Over the last decade, grape cultivation in China has rapidly progressed in terms of the cultivation area, yield, and quality, in addition to management technology. According to the statistics of the Ministry of Agriculture, by the end of 2015, the area of grape cultivation in China had reached 799,000 Ha, production was 13.669 million tonnes, and wine production was 1.14 million tonnes [1]. China's table grape production increased by 591% between 2000 and 2014, which was much faster than the overall global growth of 71% over the same period. The share of the country in world production rose accordingly from 8% in 2000 to 34% in 2014, which made China the world's largest table grape

producer, with 9 million tones. The production of wine grapes reached 180,000 tonnes in 2014 and has doubled over the past 15 years (+112% compared with 2000) [2]. Therefore, timely and accurate information about the cultivation area and spatial distribution of vineyards would provide strong support for precisely estimating its production and predicting market performance, in addition to adjusting and optimizing the planting area at a regional or country level.

Remote sensing (RS) is the ideal technique for obtaining such information because it is fast, objective, and has a wide observation range. However, previous studies using RS data have mainly focused on mapping grape varieties [3,4], retrieving physical and biochemical parameters (e.g., leaf area index and chlorophyll) [5–10], evaluating the growing state (e.g., evaporation, water stress, and soil moisture) [9,11–13], and detecting disease [14], in addition to estimating the quality [15] and production [16] of grapes. Research on vineyard identification is quite rare, and most studies have typically been based on the data with very high spatial resolution (e.g., unmanned aerial vehicle images, UAV images) [17–22]. Among these vineyard identification studies, textural features caused by unique planting approaches are clearly reflected in UAV images, and thus contribute the most to the vineyard identification results, whereas spectral features contribute less because of the similarity, in terms of the phenological phase and canopy reflectance, with other crops [23]. This may be the reason that few studies have been conducted to identify vineyards using single-/multi-phased satellite images in which the spatial resolution is greater than that of UAV images. Within the existing research of vineyard recognition based on satellite imaging, the features used are basically employed from the literature on similar crop identification/detection studies. Spectrums in the range of visible to short wave infrared are the basic features that are commonly employed by previous studies, which justified broad-band multispectral remote sensing imagery of high spatial resolution shows potential applications for vineyard identification [3,9]. Vegetation indices derived from the spectral bands are another important feature component, which are designed to characterize different physiological or biochemical features of vineyards' canopy. For example, the normalized difference vegetation index (NDVI) was used to indicate the condition of vine leaf area index (LAI) [8], the perpendicular vegetation Index (PVI) and ratio vegetation index (RVI) were adopted to present the canopy density of the vineyard [24], and band ratios and indices like red-edge/blue and the modified soil-adjusted vegetation index (MSAVI) were used to depict the difference between vineyards and other classes [25]. Texture features are often associated with very high spatial resolution, so they are widely used in UAV imagery. Since vineyard mapping using UAV images is too costly to be applied on a large scale, dense time series RS data is expected to be the solution for large-scale and accurate vineyard mapping if the subtle differences, both in spectrum and phenological phase, between vineyards and other crops over the entire growing season/year are fully considered and exploited. Therefore, a study on identifying vineyards using time series satellite images is necessary and valuable.

With the increasing success of RS techniques and the increasing number of satellites in orbit, a huge volume of RS data with a high spatial and temporal resolution is freely accessible to users all over the world, which satisfies the prerequisite for a variety of applications and enables new applications. Abundant data sources benefit the construction of time series RS data, which can fully depict phenological features of vegetation; thus, it is important to distinguish crop type from each other. To date, time series RS data have been successfully applied to many practical applications, such as land cover classification [26,27], change detection [28,29], crop type classification [30], crop phenology detection [31,32], and crop monitoring [33–35] because they can provide informative features for subsequent classification/regression algorithms.

In the context of crop type mapping, many approaches have exploited time series RS data, and can be roughly divided into three categories: First, time series data have directly served as the input of the traditional supervised classification method (e.g., random forest (RF)), where time series data are regarded as independent features. Their temporal information contributes nothing to the classification result in these approaches [36]. Second, deep learning approaches, such as convolutional neural networks (CNNs), have been applied to identify crop types by considering original long time series

data as the only input [37–41]. Within such approaches, time series images are directly stacked, then CNNs automatically construct convolutional filters of three dimensions to extract the features to serve its class prediction layer. Third, the Fourier transform and wavelet transform (WT) have been used to perform frequency analysis and extract phenological features to complete the task of crop type identification because time series RS data contemporaneously contains temporal and amplitude information [32,42]. That is, detecting sowing and harvesting date, peak point of growth, rising and falling rate of vegetation index, then treating them as the input vector of the classification algorithm.

Feeding directly stacked image data into a traditional machine learning method and CNN ignores temporal dependencies and only uses the amplitude information of the time series. However, different crops theoretically have unique temporal profiles because they have different growth cycles. Therefore, research on extracting crop phenological features and analyzing frequency characteristics, in addition to modeling growth cycles, has been conducted to pursue more reliable and accurate recognition results. Among such studies, the Fourier transform and WT are the two most common approaches that are adopted to model crop seasonality and then extract phenological features to help the consequent identification task. Because of the complex planting structure/crop rotation and parcel size, in addition to the similarity of the spectrums between different crops, the fitted growth cycle model or the extracted phenological features may not be able to fully represent the differences between crops, and thus fail to accurately identify crop types.

As aforementioned, the researches of vineyard recognition based on satellite image failed to exploit the temporal information contained in the time series of satellite images, thus, our motivation is to propose a new approach that can make not only full use of the amplitude information of each image but also the temporal information contained in whole time series, verifying both kinds of information are contributory to the identification task. The objective of this study is to integrate the advantages of the wavelet in the analysis of time series data, the superiorities of the CNN in the image recognition field, and the strengths of time series satellite image data to distinguish vineyard from other classes and made a substantial advance in mapping vineyard distribution at a regional level. The sub-objectives are as follows: (i) to verify that applying the continuous WT (CWT) to time series RS data is a stable and effective approach to obtain more information about vineyards, particularly in the frequency domain; and (ii) to confirm that feeding the scalograms generated by the CWT into the CNN is a feasible solution for identifying vineyards, and is also better than directly feeding the original time series into a traditional supervised classification method.

The remainder of the paper is organized as follows: The dataset and study area are introduced in Section 2, and the details of the proposed approach are illustrated in Section 3. Following the result reported in Section 4, a discussion and conclusions are presented in Sections 5 and 6, respectively.

2. Dataset and Study Area

Shaanxi province belongs to the arid/semi-arid region of the Loess Plateau, which is one of the seven major grape-producing regions in China. The study area is located in the central Shaanxi province and spans an area of approximately 1000 km². Throughout the entire study area, there are seven planting conditions, in general, that appear in cropland over a full year (Table 1): crop rotations of spring corn to vegetables, winter wheat to corn, vegetables to vegetables, vineyard, peach trees, greenhouse, and other forest. Note that the forest mentioned in this paper is a collection of trees planted on cropland, including sophora japonica, apple trees, persimmon trees, pine trees, and other trees except peach trees because they have a small planting area over the study region. Among the seven planting conditions, the rotation of winter wheat to corn takes up more than 60% of the planting area of the study region; vegetables and fruit trees rank second and third, respectively. Because it is extremely difficult to recognize the types of crops planted in greenhouses without the support of other data, the study only considers vines planted in normal cropland without any materials covering its canopy, and further crop type recognition is not performed for the greenhouse category.

A field survey was conducted on June 26, 2018, and 581 geo-tagged samples were collected (shown in the right panel of Figure 1). Geo-tagged samples were used to identify the unique features of each crop type. Training/testing data for each type of crop were acquired manually and independently from the Sentinel-2 image based on the analysis of ground samples and the images (many small polygons were randomly drawn to select sample pixels using ENVI software); the details are listed in Table 1. The training/testing datasets were mainly defined in pixels because the identification process was conducted at the pixel level.

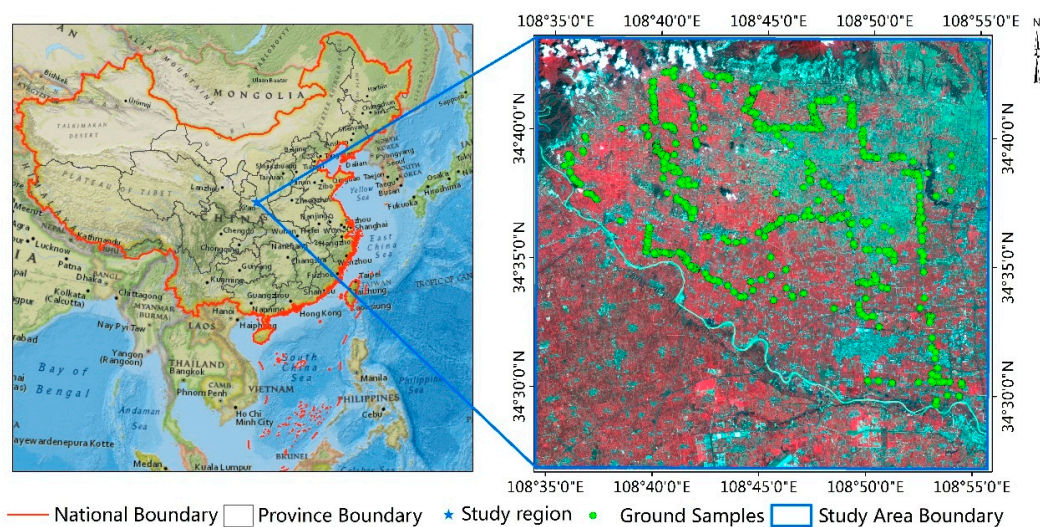


Figure 1. Geolocation of the study area.

Table 1. Summary of planting conditions and the training/testing datasets.

ID	Planting Conditions	Growing Seasons Over a Full Year	Training/Testing Dataset (Pixels)
1	Spring corn → Vegetables (SC-Veg)	2	1620/1000
2	Peach	1	2036/1000
3	Forest	1	1014/1000
4	Greenhouse (GH)	-	780/500
5	Vineyard	1	2122/1000
6	Winter wheat → Corn (WW-Cron)	2	3656/1000
7	Vegetables	3	1426/1000

As a valuable data resource for vegetation monitoring, the Sentinel-2 satellite constellation has had a revisit cycle of five days since early 2018, which makes it possible to build dense and consistent time series data all over the world to mitigate the problems of cloud and cloud shadow contamination [43]. Despite its excellent revisit ability, Sentinel-2 also performs well in spatial and spectral resolutions, including three narrow bands for cloud screening and atmospheric correction at 60 m, and three red-edge bands and two shortwave infrared bands that provide key information about vegetation at 20 m, in addition to four classical bands (i.e., blue, green, red, and near infrared bands) at 10 m. The wide swath of Sentinel-2, coupled with its five-day revisit cycle, creates the opportunity to address the challenges that remain for precise mapping and monitoring in the agricultural field.

Sentinel-2 was the only satellite data source used in the study, and the study area was fully covered by the tile of 49SBU in the Worldwide Reference System (WRS-2). The Level-1C products of Sentinel-2A/B provided by the European Space Agency were downloaded from Sentinel Hub (<https://apps.sentinel-hub.com/eo-browser/>) for the period of January 1–December 31, 2018. Finally, 33 scenes of Sentinel-2 images that had a cloud percentage lower than 50% or for which the spatial distribution of clouds did not obscure the study area were acquired to build an original image time

series (as shown in Figure 2). Atmospheric correction was performed using the Sen2Cor tool, which is a processor for Sentinel-2 Level-2A product generation and formatting. Following atmospheric correction, three commonly used vegetation indices (VIs) were calculated for each of the original time series because they have been confirmed to be highly related to the growth cycle and physical features of crops. The three VIs are normalized difference vegetation index (NDVI) [44], green chlorophyll vegetation index (GCVI) [45], and enhanced vegetation index (EVI) [46], and their formulas are as follows:

$$\text{NDVI} = \frac{\text{Nir} - \text{Red}}{\text{Nir} + \text{Red}} \quad (1)$$

$$\text{GCVI} = \frac{\text{Nir}}{\text{Green}} - 1, \quad (2)$$

$$\text{EVI} = \frac{2.5 * (\text{Nir} - \text{Red})}{(\text{Nir} + 6.0 * \text{Red} - 7.5 * \text{Blue} + 1)}, \quad (3)$$

where *Nir*, *Red*, *Green*, *Blue* is the spectral reflectance of the band, respectively. NDVI is usually interpreted as a useful indicator of LAI and photosynthetic capacity of vegetation; however, a saturation phenomenon appeared at high leaf area biomass. The development of GCVI and EVI is largely aimed at solving the saturation problem in NDVI. Additionally, GCVI has been found to have the most linear relationship with leaf area index (LAI) when compared with other VIs, and EVI is capable of reducing the influence of some atmospheric effects by including blue bands and is proportional to the vegetation biomass. These three VIs were selected to express the variation trend of each kind of vegetation among a full year in the aspects of photosynthetic capacity, vegetation biomass, and LAI, and then reveal the difference between a vineyard and other classes.

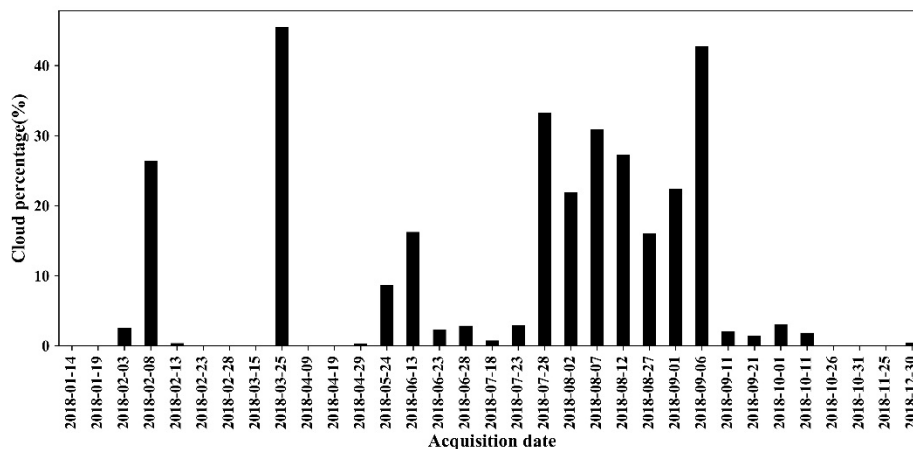


Figure 2. Cloud percentage and acquisition date of the original images.

3. Methods

As time series data can be readily built, it is beneficial for characterizing potential patterns of crops among their growing cycles, which are important for addressing the existing challenges in crop identification or completing a more complicated identification task, for example, identifying vineyards. In this study, the approach proposed to identify vineyards with time series RS data has three major parts (Figure 3): reconstructing daily time series for three VIs, applying the CWT to the reconstructed time series of the VI to generate the corresponding scalogram, and using a CNN-based classifier to identify vineyard by considering the scalograms as input. Applying the CWT to time series data allows the full exploration of the frequency information of different crops during the growing cycle in the frequency domain while simultaneously maintaining temporal information, and the scalogram is the direct carrier of these two types of information. The deep CNN is used to detect the specific pattern hidden in the scalograms to achieve the goal of identifying vineyards.

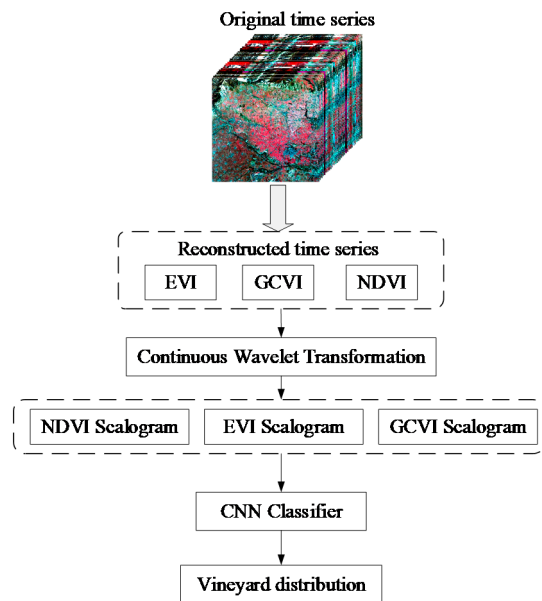


Figure 3. Technical flow chart of the proposed approach.

3.1. Iterated Savitzky-Golay Filter-Based VI Reconstruction

Since VI products have been widely used in the RS community as time series data, many more differences in the growing patterns of different crops have been expressed, and such patterns are significantly important for distinguishing crops from each other. However, because of both the satellite observation plan and the influence of clouds and poor atmospheric conditions, it is unrealistic to build a daily RS data series (e.g., NDVI time series) without interpolation or any reconstruction process. Applying a reconstruction approach to the original observation time series data is required in some operational practice for two reasons: denoising and filling missing values. By contrast, to facilitate the consequent CWT analysis, our approach requires that the input of the CWT is a daily time series for a full year. Therefore, three original VIs are reconstructed by applying an iterated Savitzky-Golay (S-G) filter-based method [47], which was proposed to provide high-quality NDVI time series data, to meet the requirement of the CWT analysis. The S-G filter method was originally proposed by Savitzky and Golay in 1964 [48], which has widely been used in data smoothing and denoising. It is a filtering method based on local polynomial least squares fitting in the time domain, and its biggest advantage is that it can ensure the shape and width of the signal while filtering out noise. Within the reconstruction approach, those positions that have an NDVI increase greater than 0.3 during 20 days are rejected and replaced by a null value before the iterated reconstruction process because such an increase cannot be caused by natural vegetation changes. As different pixels have different situations, in the most severe case, only 3 samples were set to null after applying this threshold. In addition to the reconstruction of the NDVI time series, the positions that are first replaced by a null value in the original NDVI time series are also replaced by a null value when the EVI and GCVI time series are reconstructed. During the reconstruction process, the width of the smoothing window and the degree of the smoothing polynomial are the two most important parameters; they determine the smooth extent of the reconstruction result. The wider the smoothing window, the smoother the result at sharp peaks, and the smaller the degree of the smoothing polynomial, the smoother the result, but bias may be introduced. Finally, we set the width of the smoothing window to 31 and the degree of the smoothing polynomial to 3 to obtain the optimal reconstruction result.

3.2. Continuous Wavelet Transformation

The WT has been proven to be a useful tool in the study of time series and has been applied successfully numerous times in an extraordinary range of fields because of its ability to extract various

components (e.g., seasonal, trend, and abrupt components) for time series [49]. The WT can decompose a signal directly according to the frequency, and represent it in the frequency domain distribution state in the time domain. The Fourier transform is not localized in time, whereas a wavelet is localized in time, which allows the WT to obtain time information in addition to frequency information [49]. Thus, it is a more powerful transformation for time series analysis.

A wavelet function (or wavelet, for short), is a function $\varphi \in \mathbb{R}$ with zero average (i.e., $\int_{\mathbb{R}} \varphi = 0$), normalized (i.e., $\|\varphi\| = 1$), and centered in the neighborhood of $t = 0$ [50]. Scaling φ by a positive quantity s , and translating it by $u \in \mathbb{R}$ generates a family of time-frequency atoms (i.e., the mother wavelet), $\phi_{u,s}$: [51]

$$\phi_{u,s}(t) = \frac{1}{\sqrt{s}} \phi\left(\frac{t-u}{s}\right), \quad u \in \mathbb{R}, s > 0. \quad (4)$$

Given the original signal $f \in \mathbb{R}$, the CWT of f at time u and scale s is defined as

$$Wf(u, s) = \int_{-\infty}^{+\infty} f(t) \phi_{u,s}^*(t) dt, \quad (5)$$

and it provides the frequency component (or details) of f that correspond to scale s and time location u . Wavelet coefficients $Wf(s, u)$ are obtained by continuously varying the scale and the position parameters to select different portions of the original signal and analyze different scale variations. Conversely, the original signal can be retrieved by multiplying each coefficient by the appropriate scaled and shifted wavelet. The scalogram of f is defined by the function [52]

$$\Phi(s) = \|Wf(s, u)\| = \left(\int_{-\infty}^{+\infty} |Wf(s, u)|^2 du \right)^{\frac{1}{2}}, \quad (6)$$

which represents the energy of Wf at scale s . Clearly, $\Phi(s) \geq 0$ for all scale s , and if $\Phi(s) > 0$ we say that signal f has details at scale s . Thus, the scalogram is a time-scale representation of the original signal and allows the detection of the most representative scales (or frequencies) of a signal, that is, the scales that contribute the most to the total energy of the signal. Because the term frequency is reserved for the Fourier transform, the WT is typically expressed in scales, but it is possible to convert scales to frequencies using the following equation [52]:

$$f_a = \frac{f_c}{s} \quad (7)$$

where f_a is the frequency, f_c is the central frequency of the mother wavelet, and s is the scaling factor.

Many families of wavelets exist, and they differ from each other because, for each family, a different trade-off has been made regarding how compact and smooth the wavelet appears. For example, four of the most common continuous mother wavelets were tested in this research [53]:

$$\text{Gaussian wavelet (gaus4)} : \phi(t) = 4e^{-t^2}, \quad (8)$$

$$\text{Mexican hat wavelet (mexh)} : \phi(t) = \frac{2}{\sqrt{3} \sqrt[4]{\pi}} e^{-\frac{t^2}{2}} (1 - t^2), \quad (9)$$

$$\text{Shannon wavelet (shan)} : \phi(t) = \sqrt{B} \frac{\sin(\pi B t)}{\pi B t} e^{j2\pi C t}, \quad (10)$$

where B is the bandwidth and C is the center frequency, and

$$\text{Morlet wavelet (morl)} : \phi(t) = e^{-\frac{t^2}{2}} \cos(5t). \quad (11)$$

After comparing their influence on the scalogram, we finally used the ‘‘Morlet’’ as the mother wavelet because it extracts features with equal variance in time and frequency, which ensure the

time-frequency resolution can be adapted to different signals of interest and provide a guarantee of extracting temporal features. The scale varies from one to a maximum scale, and the specific maximum scale is determined by calculating the entropy (Equation (12)) of the scalograms [54]:

$$H(X) = - \sum_{i=1}^n p_i \log p_i \quad (12)$$

where H is the entropy of X , and p_i is the probability of the i th class in X .

3.3. CNN-Based Identification

Over the last few years, CNNs have made great advances in many fields, such as reducing noise, super-resolution reconstruction [55], pan sharpening [56], image segmentation [57,58], object detection [59], change detection [60,61], and classification [62,63]. In these studies, the convolutional operation was applied in both the x and y dimensions to detect the potential architecture or patterns hidden in data. Similarly, in this study, a CNN is used to detect the distinctive characteristics of vineyards in both the time and scale dimensions in scalogram data because it has the following strengths: i) extreme versatility that allows it to approximate any type of linear or nonlinear transformation, including scaling or hard thresholding; ii) it is not necessary to design handcrafted filters, they are automatically learned by algorithms; and iii) high-speed processing because of parallel computing. In this study, a new CNN similar to LeNet [64], which combines small convolutional kernels with maximum or average pooling to learn high-level features, is built to identify vineyards. The convolution and pooling operations are conducted both on the frequency/scale and time axis in our experiment because the scalogram contains time-related frequency patterns of each planting condition. Specifically, Table 2 lists the CNN layers in order. For example, the first convolutional layer represents a convolutional kernel with three input channels, 12 output channels, and a size of 5; and the first pooling layer represents a max pooling kernel of a size of 4 and strides of 2. The activation functions for all convolutional layers and the first dense layer is an ReLU, whereas it is softmax for the last dense layer. The output of the last convolutional layer is concatenated into one vector, then fed into the fully connected layer with 100 units. Finally, there is a softmax layer with two units, which indicate non-vineyard and vineyard. To test an unknown pixel, the CNN takes its three VIs' (i.e., NDVI, EVI, and GCVI) scalograms, concatenated as a three-dimensional array, as the input, and the maximum unit of the last softmax layer is the result. For each scalogram, the size of x dimension is equal to the length of the reconstructed time series (i.e., 365), and the size of y dimension is dependent on the value of maximum scale of CWT (i.e., 200 in this research), which is determined by calculating the optimal entropy of scalogram. As three stacked scalograms are directly fed into the CNN without splitting them into several small patches, the input size of CNN is $365 \times 200 \times 3$. The construction of CNN is implemented with the assistance of Keras, and TensorFlow is used as the backend of Keras in the research.

Table 2. Convolutional neural network (CNN) architecture.

Layer (Type)	Output Shape	Kernel Size	Param #
Input Layer	(None, 200, 365, 3)	-	0
Conv2D	(None, 196, 361, 12)	(5,5)	912
Max_Pooling2D	(None, 97, 179, 12)	(4,4)	0
Conv2D	(None, 93, 175, 24)	(5,5)	7224
Max_Pooling2D	(None, 46, 87, 24)	(2,2)	0
Conv2D	(None, 42, 83, 48)	(5,5)	28,848
Max_Pooling2D	(None, 21, 41, 48)	(2,2)	0
Flatten	(None, 41328)	-	0
Dense	(None, 100)	-	4,132,900
Dense	(None, 2)	-	202

Total params: 4,170,086; Trainable params: 4,170,086; Non-trainable params: 0.

3.4. Experiment Settings and Accuracy Assessment

In addition to the proposed approach, RF [65], one of the most commonly used and effective methods for mapping land cover and crop types using RS data, was selected to identify vineyards under the same circumstances to serve as a control group, because such applications often exploit multi-phased images while RF is not able to capture temporal information contained in multi-phased images. The original time series of three VIs were directly concatenated as one vector, and then the fitted RF classifier was applied to perform binary classification. The RF algorithm is implemented with the assistance of the Scikit-learn library, which is a machine learning library in Python. Within the fitted RF model, 50 trees were used, the maximum depth of each tree is automatically determined by the algorithm, and the maximum number of the features considered to be used when splitting each node during the construction of a tree is set to the square root of the number of the total features (i.e., 10 in this research), other parameters were set to default values. The input of RF is a vector of 99 elements (33 timestamps \times 3 indices). Note that the same training and testing datasets listed in Table 1 were used to train the RF classifier and test its performance. Finally, confusion matrices that include the producer's accuracy (Pro.'s Acc.), user's accuracy (Usr.'s Acc.), and overall accuracy (OA) were used to assess the identification performance.

4. Results

4.1. Reconstructed Time Series of the Three VIs

Noise and oscillations remained in the original time series of NDVI, EVI, and GCVI, which prevented them from illustrating the general and real growing trend of different crops over a full year. An appropriate reconstruction approach could rebuild a daily time series while removing noise and filling missing values to reveal the general trends of different crops and identify the differences among them. Figure 4 shows examples of the reconstruction results of the three VIs for seven typical crops or crop rotations that were summarized in Section 2 during a full year throughout the entire study region. Generally, they all achieved good reconstructed results in terms of R^2 (as shown in Table 3). For crop rotations of spring corn to vegetables and winter wheat to corn (Figure 4a,f), the reconstructed results exactly depicted the phenological characteristics of the two crops and exhibited a stable development trend in the NDVI, EVI, and GCVI time series. The group of peach trees, forest, and vineyards (Figure 4b,c,e) all had unimodal development curves in the three VIs because there was no rotation or changes in the middle year. By contrast, they had different rising/decline rates and smoothness of peaks because of their unique features, such as flowering and deciduous times, planting structure and density, and plant morphology. Greenhouse had relatively complex development curves for all three VIs' reconstructed results over a full year (Figure 4d), and the reasons can roughly be summarized in three categories: (i) the insulation materials that covered the outside of the greenhouse were different; (ii) numerous types of crops were planted in the greenhouse, and each crop had a different canopy reflectance; (iii) the greenhouse opposed the faster and unfixed phenological phase because it provided a superior temperature and humidity environment. The developing curve of vegetables planted on normal cropland generally had three peaks because vegetable farmers in our study region typically grow three seasons of vegetables, and the three reconstructed time series successfully characterized this feature (Figure 4g). Generally, the iterated S-G reconstruction approach produced the projected daily time series data of the three VIs for all planting conditions except the greenhouse, but the substandard reconstructed results of the greenhouse only demonstrated that its reconstruction was subject to the aforementioned factors rather than the unsuitableness of the reconstruction approach.

From the view of the overall development trend of the reconstructed time series of the seven planting conditions (Figure 4), it was possible to distinguish the vineyard from crop rotations of winter wheat to corn and spring corn to vegetables, greenhouse, and vegetables according to the number of growing peaks during the entire year; the vineyard had only growing peak, but the other four planting

conditions had more than one. Although peach trees, forest, and the vineyard all had only one growing peak during the entire year, some differences still existed in the reconstructed time series of the VIs. For example, the vineyard developed slower than peach trees and forest during March and April in the NDVI case, and in the EVI case, the vineyard had the flattest development trend, and its values were lower than peach trees and forest after June. Both the development trend and the variations of the VIs values over the entire year made it possible to distinguish the vineyard from peach trees and forest.

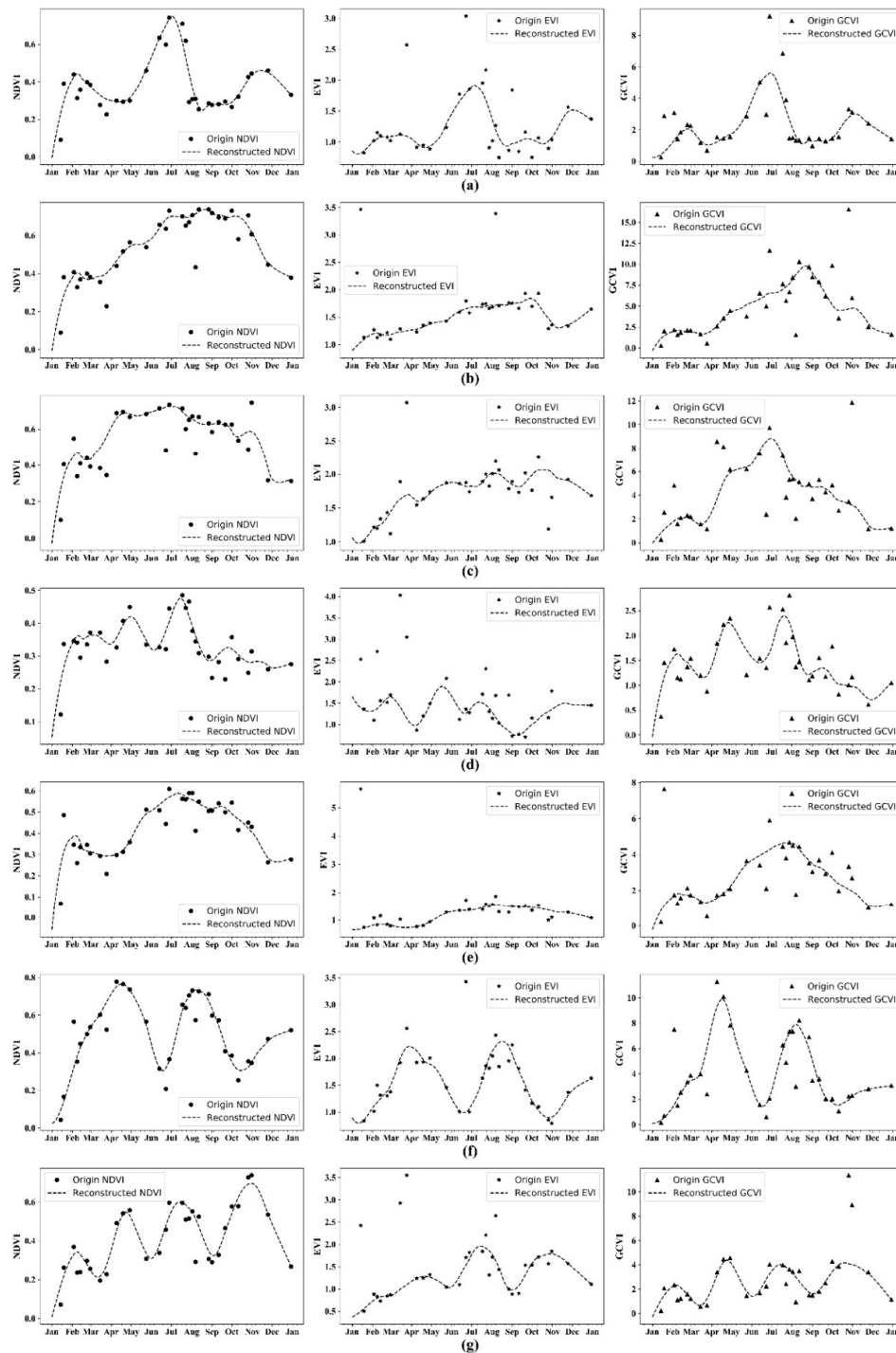


Figure 4. Example of the vegetation indices' (VIs) time series reconstruction results for (a) spring corn-vegetables, (b) peach trees, (c) forest, (d) greenhouse, (e) vineyard, (f) winter wheat-corn, and (g) vegetables.

Table 3. Reconstruction accuracy of example time series in Figure 4 (R^2).

	NDVI	EVI	GCVI
Spring corn-Vegetables	0.825	0.648	0.716
Peach trees	0.812	0.523	0.526
Forest	0.665	0.634	0.343
Green house	0.726	0.216	0.652
Vineyard	0.752	0.795	0.311
Winter wheat-Corn	0.875	0.559	0.660
Vegetables	0.862	0.486	0.559

4.2. Scalograms

To determine the optimal maximum value of the scale, the maximum scale was varied from 1 to 500 to explore the relationship between the maximum scale and the entropy of the scalogram for different planting conditions. Figure 5 shows how the scalograms' entropy relates to the increase of the maximum scale in different cases. Clearly, either three VIs or seven planting conditions demonstrate a uniform trend; that is, as in maximum scale increased, the entropy of the scalogram increased accordingly. However, an entropy saturation phenomenon appeared gradually in three VIs and seven planting conditions when the maximum scale value was greater than 200, particularly in the NDVI case. Therefore, the scales were set to the range of 1–200 in the study, and the consequent CWTs were performed based on this scale setting.

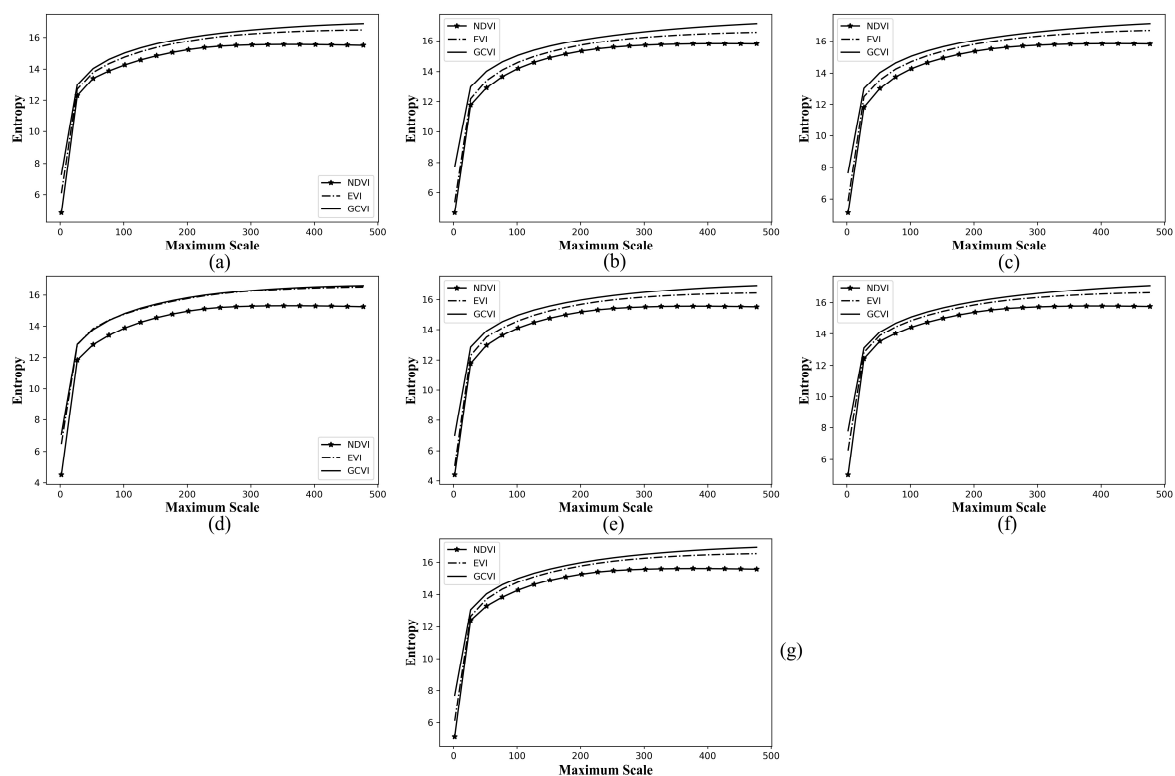


Figure 5. Relation between the entropy of the scalogram and maximum scale for (a) spring corn-vegetables, (b) peach trees, (c) forest, (d) greenhouse, (e) vineyard, (f) winter wheat-corn, and (g) vegetables.

Once the mother wavelet and scales were determined, applying the specific CWT to the reconstructed time series yielded the corresponding scalogram that represented the time series data in both the frequency and time domain. As the scalogram examples in Figure 6 show, in the two-dimensional feature space, that is, frequency and time, each planting condition had a different

signature in the three VIs' cases. For example, the crop rotations of spring corn to vegetables (Figure 6a) and winter wheat to corn (Figure 6f) differed in both the coefficient distribution pattern and its magnitude; particularly in the GCVI case, they had great dissimilarities. Similarly, noticeable differences in all three scalograms easily distinguished the vineyard (Figure 6e) from peach trees (Figure 6b) and other forest (Figure 6c), even though many subtle differences still existed among the latter two classes, such as the magnitude difference in the EVI scalograms and the difference in the distribution shape in the GCVI scalograms. For the greenhouse (Figure 6d) and vegetables (Figure 6g), both the magnitude difference of the coefficient and the difference in the distribution pattern in all three scalograms provided sufficient evidence to distinguish them from each other. From the view of distinguishing the vineyard from the other six classes, although NDVI scalograms could not provide an obvious visual difference to distinguish vineyard from crop rotation of spring corn to vegetables, greenhouse, and vegetables, the distribution shape and magnitude differences in the EVI and GCVI scalograms excluded the rotation of spring corn to vegetables, greenhouse, and vegetables. To summarize, the magnitude differences and distribution patterns in all three types of scalograms ensured that time-related frequency features provided sufficient detail to the CNN classifier to successfully complete the vineyard identification task.

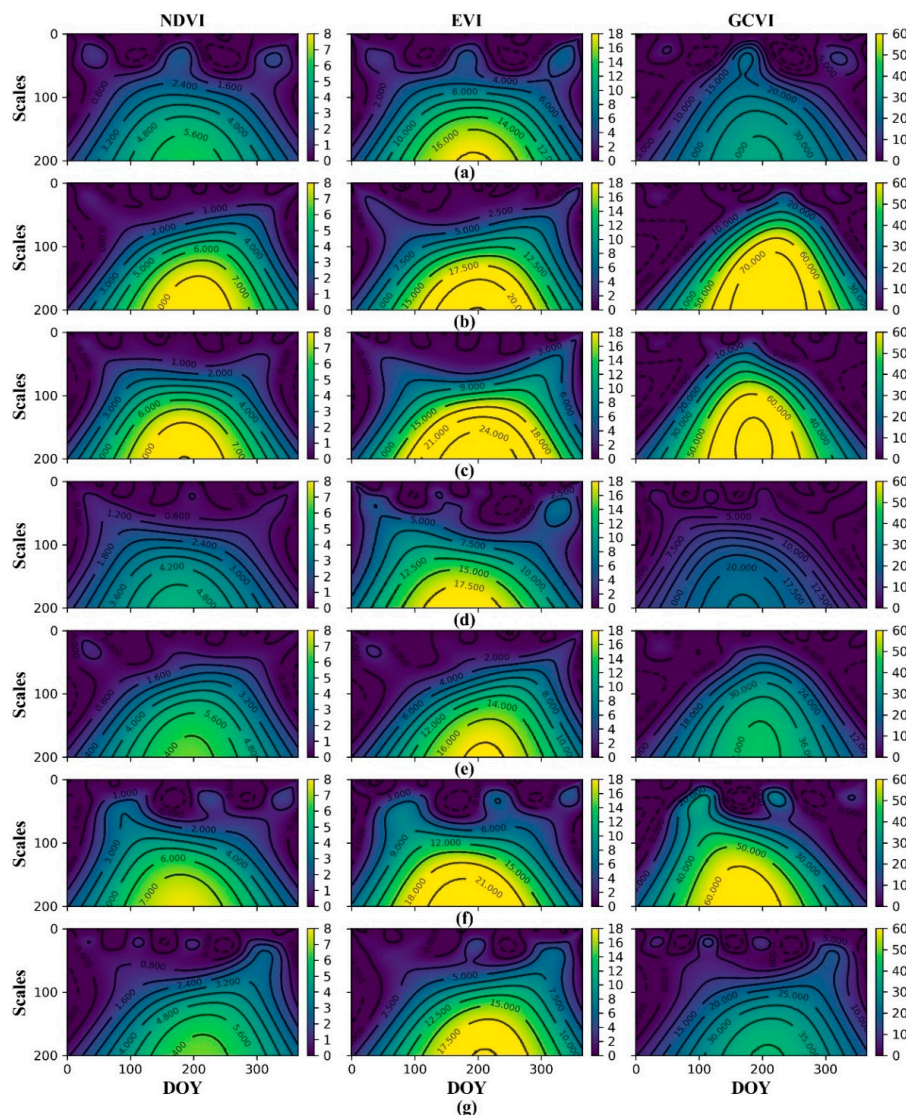


Figure 6. Examples of scalograms for (a) spring corn-vegetables, (b) peach trees, (c) forest, (d) greenhouse, (e) vineyard, (f) winter wheat-corn, and (g) vegetables.

4.3. CNN Classification Results

Considering the stacked scalograms of NDVI, EVI, and GCVI as the input of the CNN, with the learning processes of three convolution layers, three max pooling layers, and a dense layer, the last softmax layer of the CNN indicated the identification result. If the value of the first unit was greater than that of the second unit, then this denoted a non-vineyard; conversely, it denoted a vineyard. Figure 7b represents the identification results of the proposed approach, from which we determined that, in the entire study region, most vines were located in the northwest, with less planting in the northeast and southeast. Figure 7c illustrates the identification results of RF, which share a similar trend with the CNN in the overall distribution pattern. However, RF misclassified many non-vineyards as vineyards and omitted some vineyards that had high similarity with other classes (e.g., peach tree), and the result was that its total planting acreage of the vineyard was more than that of the CNN by 29.8% (Figure 7d). Furthermore, the confusion matrices shown in Table 4 quantitatively characterized the difference between the CNN and RF in vineyard identification. The overall accuracy of the CNN was 89.66%, which was 9.87% higher than that of RF (79.78%). In addition to the overall accuracy, the CNN was also superior to RF in terms of producer's accuracy and user's accuracy. For example, the producer's accuracy of vineyard was 92.90% for the CNN and 60.60% for RF, which was in accordance with RF's higher omission rate and further proved that the CNN had a better capability than RF to distinguish the vineyard from other classes, even if it was very similar to other classes. Note that the user's accuracy of the vineyard was far lower than that of the non-vineyard for both the CNN and RF. This is largely because of the imbalance of the testing sample size (5500 pixels for non-vineyard and 1000 pixels for vineyard). Although the user's accuracies of vineyard and non-vineyard seemed to be less comparable, the huge difference of 21% in the user's accuracy of vineyard between the CNN and RF verified that RF misclassified more non-vineyards as vineyards than the CNN. To summarize, both the spatial distribution map and confusion matrices proved that the approach that combined the WT and CNN was comprehensively superior to the approach that combined RF and the original time series in terms of identifying the vineyard using the Sentinel-2 time series.

Table 4. Confusion matrices for the CNN and random forest (RF).

		Prediction					
		CNN			RF		
		Non-vineyard	Vineyard	Pro's Acc.	Non-vineyard	vineyard	Pro's Acc.
Truth	Non-vineyard	4899	601	89.07%	4580	920	83.27%
	Vineyard	71	929	92.90%	394	606	60.60%
	Usr's Acc.	98.57%	60.72%	89.66%	92.08%	39.71%	79.78%

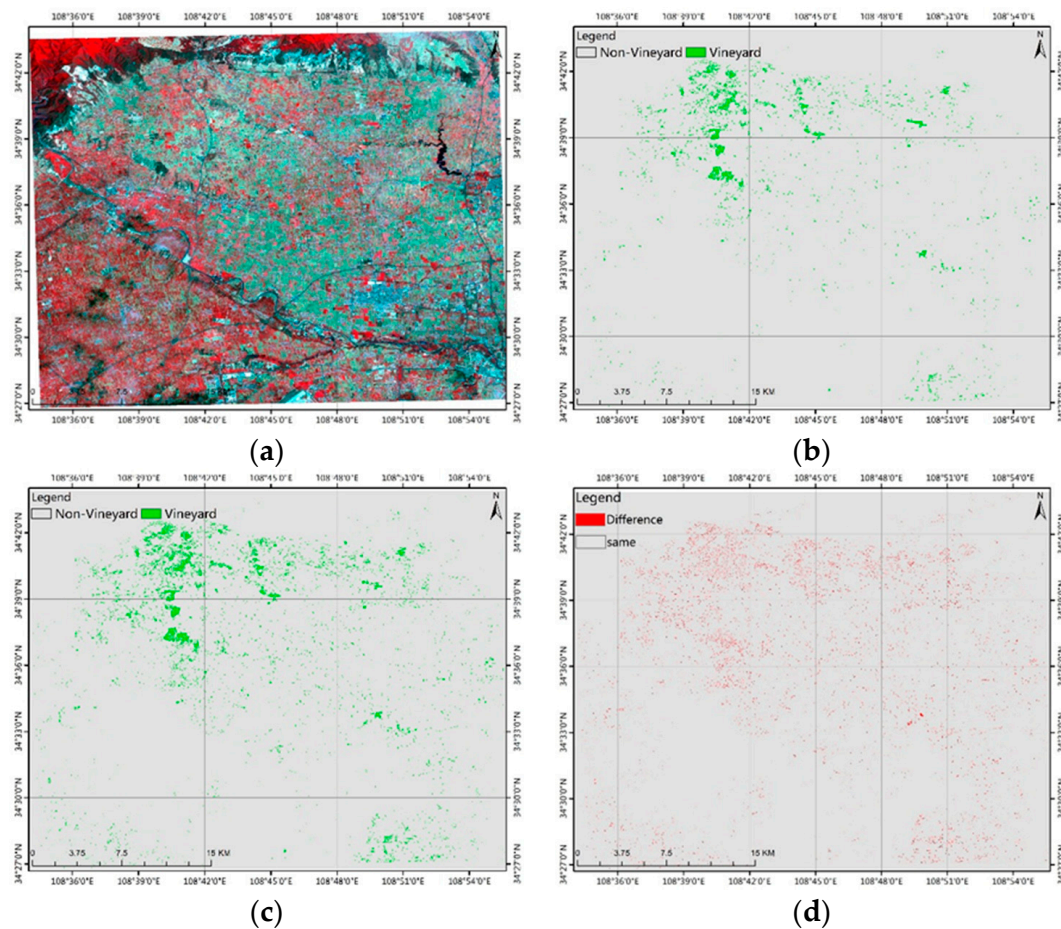


Figure 7. Vineyard identification comparison: (a) original image, (b) result of the CNN, (c) result of RF, and (d) difference between CNN and RF.

5. Discussion

5.1. Variation of the Entropy of the Time Series Data

In this study, Sentinel-2 images with a cloud-cover-percentage lower than 50% were collected, and NDVI, EVI, and GCVI were then calculated as the original time series. Based on the iterated S-G approach, three time series were reconstructed and then served as the input of the CWT, which produced scalograms to present the time-related frequency information of the reconstructed time series. Therefore, exploring how the reconstruction process and CWT affects the amount of information and whether they are class sensitive is necessary for the further application of the proposed approach. The information entropy was used as the only criterion throughout the assessment process.

For each VI, the average entropies of each class in the original, reconstructed, and scalogram data were calculated, as shown in Figure 8. Clearly, either the three VIs or seven types of planting conditions indicated a uniform trend of the entropy variation across the original, reconstructed, and scalogram data; that is, the entropy value was approximately 5, 8.5, and 15.5 for the original, reconstructed, and scalogram data, respectively. Thus, the reconstruction process and CWT approximately increased the information amount by 70% and 210%, respectively. Most importantly, such a variation of the entropy was not sensitive to the class, and the trend remained the same over the three VIs. The reason for this phenomenon is largely because the three VIs were calculated based on the fixed and same image collection, and the reconstruction approach was not sensitive to the type of VI. This means that once the original time series is fixed, the reconstruction approach and CWT will exert similar effects across different classes and VIs.

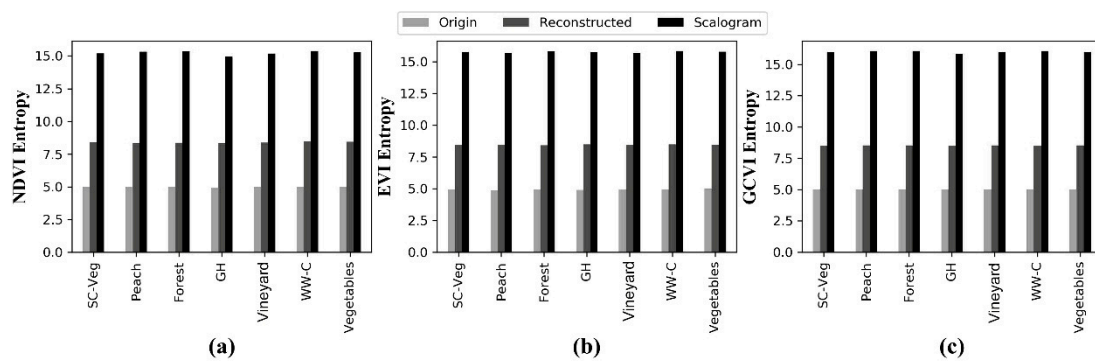


Figure 8. Comparison of the entropy variation in (a) the normalized difference vegetation index (NDVI), (b) enhanced vegetation index (EVI), and (c) green chlorophyll vegetation index (GCVI).

5.2. How the Mother Wavelet and Scale Range Influence the Entropy

As mentioned in Section 3.2, there are many families (types) of mother wavelets. They differ from each other because each type of wavelet has a different shape, smoothness, compactness, and is useful for different purposes. To assess the influence of different mother wavelets on the reconstructed NDVI, EVI, and GCVI time series.

For each mother wavelet, the scales varied from one to a maximum scale, and the information entropy continued to be used as the criteria to assess the influence. Figure 9 illustrates the results of all four mother wavelets in the cases of the three VI and seven planting conditions; note that the maximum scale was iterated from 2 to 502 with an interval of 25.

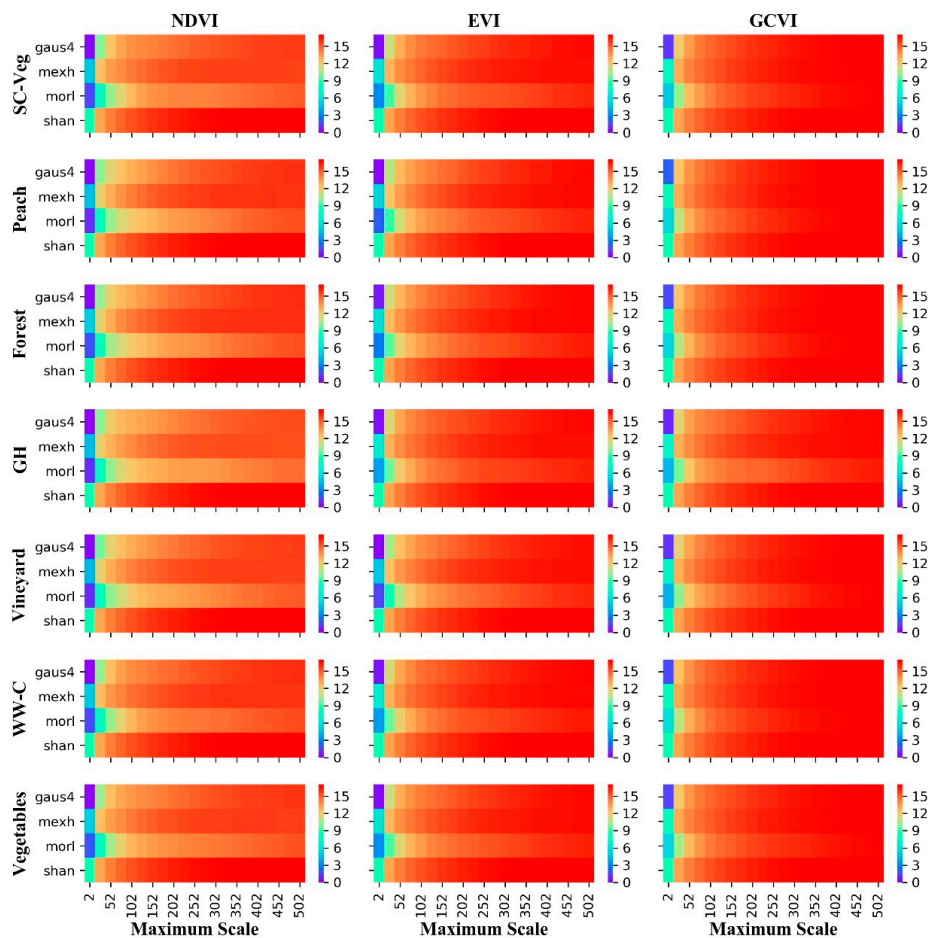


Figure 9. Varied entropy of scalograms.

Clearly, with the increase of the maximum scale, all the mother wavelets performed similarly in all cases; that is, the saturation phenomenon in the entropy also appeared in the *gaus*, *shan*, and *morl* cases when the maximum scale was greater than 200. *Mexh* and *shan* had a similar speed to *gaus* and *morl* in reaching the saturation point, whereas their entropies were slightly lower than that of the latter two, particularly in the NDVI case. However, the differences between the four mother wavelets were not significant, particularly in the EVI and GCVI cases. From the view of planting conditions, for all three VIs, the four mother wavelets also demonstrated similar and uniform trends in different planting conditions, which further proved that the mother wavelet was not sensitive to the class. Therefore, it was safe to set *Morlet* as the mother wavelet to apply the CWT to the reconstructed time series.

5.3. Difference Between the Proposed Approach and the Traditional Approach

Because the experimental results shown in Section 4.3 proved that using the CWT and a CNN achieved better performance than using RF and the original time series in identifying the vineyard with the Sentinel-2 time series, it is meaningful to discuss the difference between these two approaches and the corresponding reasons for the purpose of further study and application. Feeding RF with directly stacked original, multi-temporal images results in two adverse outcomes: (i) the direct stacking process of multi-temporal images lost time information and RF considered each feature as independent rather than time-related; and (ii) the same or similar amplitude of features at different times enhanced the redundancy problem, which had a negative impact on the RF classifier. Additionally, the similarity between the “spectrums” of the original time series confused RF in terms of distinguishing vineyards from other crops. For example, vineyards with a higher canopy cover were similar to peach trees, so RF misclassified them as non-vineyard. Conversely, applying the CWT to the reconstructed time series in the time-frequency domain better observed the local characteristics of the original time series and simultaneously observed the time and frequency information. As a result of the CWT, the scalogram represented time-related frequency (scale) information and formed a unique distribution pattern for each type of planting condition (Figure 6). The convolution and max pooling process of the CNN captured the difference of the distribution pattern in the scalogram and then distinguished the vineyard from other classes. Compared with the traditional approach, the proposed approach has two advantages: (i) time information that remained in the time series of the VIs was reserved and used in the identification process; (ii) the excellent ability of the CNN in image identification was fully exploited. Therefore, as confirmed by this study, the proposed approach was comprehensively superior to the traditional approach in terms of identifying vineyards using the Sentinel-2 time series.

5.4. Difference Between Proposed Approach and Long Short-Term Memory Networks

A Long Short-Term Memory (LSTM) network is a special type of Recurrent Neural Network (RNN) and has been widely used to analyze time series data due to their ability to capture long-term dependencies. Despite the abundant successful applications in other time series scenarios like voice recognition, machine translation, RNNs have also achieved well performance in the remote sensing community, for example, land cover classification [66–68], crop identification [69], change detection [70]. Since we have already reconstructed the original time series of three VIs so as to obtain three stable full-year time series in the proposed approach, which also makes it possible to apply an LSTM model to identify vineyards based on these three reconstructed time series. Therefore, an LSTM model, the structure of which was summarized in Table 5, was established to test the difference between itself and the proposed approach.

The construction environment, deep learning framework, and data used for training/testing of the LSTM model are consistent with the CNN approach. Note that three full-year time series vegetation indices were simultaneously used, so, the input shape of the LSTM was 356×3 (i.e., 365 days \times 3 VIs), and only the last neural of LSTM returned a sequence to serve as the input of last dense layer; it only has one output neural and was designed to do binary classification (i.e., Vineyard or Non-Vineyard).

Table 5. LSTM structure.

Layer Type	Input/Output Shape	Param #
Input layer	(365,3)	0
Lstm_1 (LSTM)	(None,100)	41,600
Dense_1 (Dense)	(None,1)	101
Total params: 41,701; Trainable params: 41,701; Non-trainable params: 0		

The LSTM model achieved an overall accuracy of 85.28%, and the detail of its identification map is illustrated in Figure 10. LSTM has a similar performance in terms of OA with respect to the approach of CWT+CNN, however, the fact that the user's accuracy (51.24%) is lower than that of the CNN approach (60.72%) indicated that LSTM has misclassified more non-vineyard pixels into the vineyard class, which is consistent with the final distribution map (Figure 4b, Figure 10). In general, the performance of LSTM is slightly inferior to the CWT+CNN but much better than the RF model. It demonstrated that both CW+CNN and LSTM had captured time-dependent features of long time series data. Regarding the difference in identification accuracy between CWT+CNN and LSTM, the potential reasons may include two aspects: (i) CWT and LSTM have totally different principles in extracting temporal information in long time series data; (ii) CNN and LSTM have different structures and different numbers of trainable parameters, but there is no doubt that both methods captured and made full use of the time-dependent information in the long time series data in our vineyard identification research.

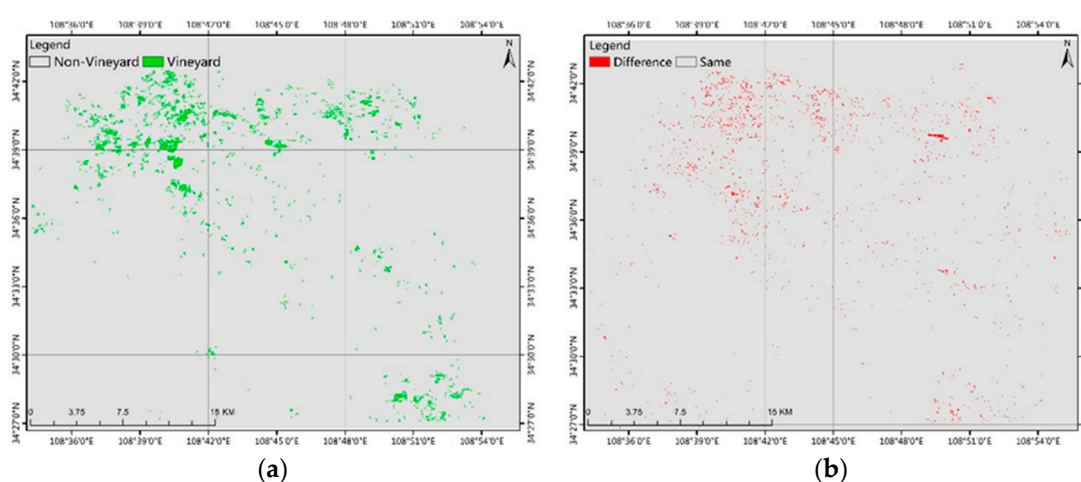


Figure 10. (a) Vineyard identification result of the LSTM model, (b) difference between CNN and LSTM.

5.5. Applicability in Other Practical Scenarios

The cost of building a time series of RS data has decreased because an increasing number of satellites have been launched into space, and the volume of usable data has largely increased. The time series observation is more capable of characterizing the features of the changes of land surface over a long time, which provides researchers with the opportunity to conduct numerous studies, such as phenological phase detection, crop growing state monitoring, land use/landcover change detection, and crop type identification. In this study, the time series of three VIs derived from Sentinel-2 data were used to identify vineyards based on the combination of the CWT and a CNN. For the CWT, we observed that the scalogram, which was the result of applying the CWT to a reconstructed time series, contained more abundant information than the original and reconstructed time series because it simultaneously reserved the frequency and time information. More importantly, the CWT was not sensitive to the mother wavelet and ground classes when it was applied to NDVI, EVI, and GCVI time series, which implies that the CWT is stable and can be used in other frequency analysis scenarios.

In addition to the CWT, the CNN was another important part of the proposed approach. As the CNN has achieved tremendous success in the fields of object detection, image classification, and handwriting recognition, it has become a common, effective solution to the image classification task because of its unique properties. In this study, the CNN also achieved a better identification result than RF. All the previous studies conducted to classify images using CNN proves that the CNN is suitable for image classification. Therefore, the proposed approach is considered to be applicable to other applications based on time series RS data (e.g., crop type classification, mapping land use/landcover). One drawback of the proposed approach is that the computational cost is higher than that of using the original time series data and a traditional supervised method (e.g., RF) to perform the classification task. The two classification processes were conducted on a Think Station P710 with 120 GB of memory and 2 Intel(R) Xeon(R) CPU E5-2640 (2.4 GHz, 40 cores); RF consumed 1.2 min, and our approach consumed 5.6 min (only including generating scalograms and classification with CNN). The time series construction of the whole study region consumed 87 min. However, with the overall improvement of computational power and the decline of computational cost, it appears that the relatively higher computational cost of the complex solution will not be a problem in the future.

6. Conclusions

Grape, either table grape or wine grape, is widely cultivated in China because of its high economic value. In this study, an approach was designed to identify a vineyard using time series RS data, including original time series reconstruction, the CWT of the reconstructed time series, and training a CNN classifier. First, three VIs (i.e., NDVI, EVI, and GCVI) were derived from the original Sentinel-2 image collection to form the original time series; second, an iterated S-G method that focused on removing outliers and filling missing values was applied to the original time series to yield a daily time series over a full year; third, the CWT was applied to each previous reconstructed time series to generate the corresponding scalogram; and finally, three scalograms were stacked together to serve as the input of the CNN classifier to complete the identification task. The results demonstrated the following: (i) the CWT had stable performance on the reconstructed time series, which provided the CNN classifier with more time-related frequency information, and was not sensitive to the mother wavelet and ground classes; (ii) the CNN classifier with scalograms as input achieved a better vineyard identification result than RF, with an overall accuracy of 89.66% (overall accuracy was 79.78%), thereby making great progress by improving the overall accuracy by 9.87%; and (iii) the CNN captured the unique and subtle features that hide in the scalogram to distinguish the vineyard from other classes, and the combination of the CWT and CNN has more accurate performance than RF in vineyard identification using time series satellite images.

To the best of our knowledge, a few studies have been conducted to identify vineyards using the time series of satellite image data; thus, this is the first time that the proposed approach has been applied to vineyard identification. The proposed approach fully explored the frequency information of time series data by applying the CWT, and then used the advantages of the CNN in image recognition to capture those distinguishing features in scalograms. Since the result has shown that applying CWT to a time series of vegetation indices, in terms of the effectiveness in extracting information of time series, is insensitive to the type of vegetation index, surface category, and type of mother wavelet, the information entropy is stably increased after applying CWT to time series. Furthermore, just as the CNN has achieved good results in classifying scalogram images in our research, many previous studies on the classification of images using CNN have also proved that CNN has unique advantages for image recognition. Therefore, the approach is considered to be able to be applied to many other practical scenarios, and further researches on testing the performance of the proposed approach in applications like classifying other crops (e.g., corn, soybean, and rice) and mapping land use/landcover based on time series images are recommended and appreciated.

Author Contributions: Conceptualization, L.Z. and Q.L.; methodology, L.Z., Y.Z., and H.W.; software, L.Z.; validation, L.Z. and H.W.; formal analysis, L.Z.; investigation, L.Z. and H.W.; writing—original draft preparation, L.Z. and Y.Z.; writing—review and editing, L.Z.; visualization, L.Z.; funding acquisition, Q.L. and X.D.

Funding: This research was funded by the National Key R&D Program on Monitoring, Early warning and Prevention of Major National Disaster grant number 2017YFC1502802, the National Science Foundation of China grant number 41701486, and Open Fund of Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions (Henan University), Ministry of Education grant number GTYR201808.

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

References

1. National Bureau of Statistics of China. *China Statistical Yearbook*; National Bureau of Statistics of China: Beijing, China, 2016.
2. *Table and Dried Grapes. Fao-iv Focus 2016: Non-Alcoholic Products of the Vitivinicultural Sector Intended for Human Consumption: Statistical Report*; FAO: Italy, Rome, 2016.
3. Lacar, F.M.; Lewis, M.M.; Grierson, I.T. Use of hyperspectral imagery for mapping grape varieties in the barossa valley, South Australia. In Proceedings of the IEEE 2001 International Geoscience and Remote Sensing Symposium, Sydney, Australia, 9–13 July 2001; pp. 2875–2877.
4. Khaliq, A.; Comba, L.; Biglia, A.; Aimonino, D.R.; Chiaberge, M.; Gay, P. Comparison of satellite and uav-based multispectral imagery for vineyard variability assessment. *Remote Sens.* **2019**, *11*, 436. [[CrossRef](#)]
5. Sun, L.; Gao, F.; Anderson, M.C.; Kustas, W.P.; Alsina, M.M.; Sanchez, L.; Sams, B.; McKee, L.; Dulaney, W.; White, W.A.; et al. Daily mapping of 30 m lai and ndvi for grape yield prediction in california vineyards. *Remote Sens.* **2017**, *9*, 317. [[CrossRef](#)]
6. Rey-Carames, C.; Diago, M.P.; Martin, M.P.; Lobo, A.; Tardaguila, J. Using rpas multi-spectral imagery to characterise vigour, leaf development, yield components and berry composition variability within a vineyard. *Remote Sens.* **2015**, *7*, 14458–14481. [[CrossRef](#)]
7. Lamb, D.W.; Weedon, M.M.; Bramley, R.G.V. Using remote sensing to predict grape phenolics and colour at harvest in a cabernet sauvignon vineyard: Timing observations against vine phenology and optimising image resolution. *Aust. J. Grape Wine Res.* **2004**, *10*, 46–54. [[CrossRef](#)]
8. Johnson, L.F.; Roczen, D.E.; Youkhana, S.K.; Nemani, R.R.; Bosch, D.F. Mapping vineyard leaf area with multispectral satellite imagery. *Comput. Electron. Agric.* **2003**, *38*, 33–44. [[CrossRef](#)]
9. Zarco-Tejada, P.J.; Berjon, A.; Lopez-Lozano, R.; Miller, J.R.; Martin, P.; Cachorro, V.; Gonzalez, M.R.; de Frutos, A. Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sens. Environ.* **2005**, *99*, 271–287. [[CrossRef](#)]
10. White, W.A.; Alsina, M.M.; Nieto, H.; McKee, L.G.; Gao, F.; Kustas, W.P. Determining a robust indirect measurement of leaf area index in california vineyards for validating remote sensing-based retrievals. *Irrig. Sci.* **2019**, *37*, 269–280. [[CrossRef](#)]
11. Matese, A.; Baraldi, R.; Berton, A.; Cesaraccio, C.; Di Gennaro, S.F.; Duce, P.; Facini, O.; Mameli, M.G.; Piga, A.; Zaldei, A. Estimation of water stress in grapevines using proximal and remote sensing methods. *Remote Sens.* **2018**, *10*, 114. [[CrossRef](#)]
12. Soliman, A.; Heck, R.J.; Brenning, A.; Brown, R.; Miller, S. Remote sensing of soil moisture in vineyards using airborne and ground-based thermal inertia data. *Remote Sens.* **2013**, *5*, 3729–3748. [[CrossRef](#)]
13. Loggenberg, K.; Strever, A.; Greyling, B.; Poona, N. Modelling water stress in a shiraz vineyard using hyperspectral imaging and machine learning. *Remote Sens.* **2018**, *10*, 202. [[CrossRef](#)]
14. Mohite, J.; Trivedi, M.; Surve, A.; Sawant, M.; Urkude, R.; Pappula, S. Hybrid classification-clustering approach for export-non export grape area mapping and health estimation using sentinel-2 satellite data. In Proceedings of the 2017 6th international conference on agro-geoinformatics, Fairfax, VA, USA, 7–10 August 2017; pp. 196–201.
15. Meggio, F.; Zarco-Tejada, P.J.; Nunez, L.C.; Sepulcre-Canto, G.; Gonzalez, M.R.; Martin, P. Grape quality assessment in vineyards affected by iron deficiency chlorosis using narrow-band physiological remote sensing indices. *Remote Sens. Environ.* **2010**, *114*, 1968–1986. [[CrossRef](#)]
16. Cunha, M.; Marcal, A.R.S.; Silva, L. Very early prediction of wine yield based on satellite data from vegetation. *Int. J. Remote Sens.* **2010**, *31*, 3125–3142. [[CrossRef](#)]

17. Fiorillo, E.; Crisci, A.; De Filippis, T.; Di Gennaro, S.F.; Di Blasi, S.; Matese, A.; Primicerio, J.; Vaccari, F.P.; Genesio, L. Airborne high-resolution images for grape classification: Changes in correlation between technological and late maturity in a sangiovese vineyard in central italy. *Aust. J. Grape Wine Res.* **2012**, *18*, 80–90. [[CrossRef](#)]
18. Delenne, C.; Rabatel, G.; Deshayes, M. An automatized frequency analysis for vine plot detection and delineation in remote sensing. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 341–345. [[CrossRef](#)]
19. Pobleto-Echeverria, C.; Olmedo, G.F.; Ingram, B.; Bardeen, M. Detection and segmentation of vine canopy in ultra-high spatial resolution rgb imagery obtained from unmanned aerial vehicle (uav): A case study in a commercial vineyard. *Remote Sens.* **2017**, *9*, 268. [[CrossRef](#)]
20. Cinat, P.; Di Gennaro, S.F.; Berton, A.; Matese, A. Comparison of unsupervised algorithms for vineyard canopy segmentation from uav multispectral images. *Remote Sens.* **2019**, *11*, 1023. [[CrossRef](#)]
21. Sertel, E.; Yay, I. Vineyard parcel identification from worldview-2 images using object-based classification model. *J. Appl. Remote Sens.* **2014**, *8*, 17. [[CrossRef](#)]
22. Wassenaar, T.; Robbez-Masson, J.M.; Andrieux, P.; Baret, F. Vineyard identification and description of spatial crop structure by per-field frequency analysis. *Int. J. Remote Sens.* **2002**, *23*, 3311–3325. [[CrossRef](#)]
23. Delenne, C.; Durrieu, S.; Rabatel, G.; Deshayes, M.; Bailly, J.S.; Lelong, C.; Couteron, P. Textural approaches for vineyard detection and characterization using very high spatial resolution remote sensing data. *Int. J. Remote Sens.* **2008**, *29*, 1153–1167. [[CrossRef](#)]
24. Dobrowski, S.Z.; Ustin, S.L.; Wolpert, J.A. Remote estimation of vine canopy density in vertically shoot-positioned vineyards: Determining optimal vegetation indices. *Aust. J. Grape Wine Res.* **2002**, *8*, 117–125. [[CrossRef](#)]
25. Karakizi, C.; Oikonomou, M.; Karantzalos, K. Vineyard detection and vine variety discrimination from very high resolution satellite data. *Remote Sens.* **2016**, *8*, 235. [[CrossRef](#)]
26. Gomez, C.; White, J.C.; Wulder, M.A. Optical remotely sensed time series data for land cover classification: A review. *ISPRS J. Photogramm. Remote Sens.* **2016**, *116*, 55–72. [[CrossRef](#)]
27. Inglada, J.; Vincent, A.; Arias, M.; Tardy, B.; Morin, D.; Rodes, I. Operational high resolution land cover map production at the country scale using satellite image time series. *Remote Sens.* **2017**, *9*, 95. [[CrossRef](#)]
28. Frolking, S.; Hagen, S.; Milliman, T.; Palace, M.; Shimbo, J.Z.; Fahnestock, M. Detection of large-scale forest canopy change in pan-tropical humid forests 2000–2009 with the seawinds ku-band scatterometer. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 2603–2617. [[CrossRef](#)]
29. Sexton, J.O.; Urban, D.L.; Donohue, M.J.; Song, C.H. Long-term land cover dynamics by multi-temporal classification across the landsat-5 record. *Remote Sens. Environ.* **2013**, *128*, 246–258. [[CrossRef](#)]
30. Vuolo, F.; Neuwirth, M.; Immitzer, M.; Atzberger, C.; Ng, W.T. How much does multi-temporal sentinel-2 data improve crop type classification? *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *72*, 122–130. [[CrossRef](#)]
31. Carrao, H.; Goncalves, P.; Caetano, M. A nonlinear harmonic model for fitting satellite image time series: Analysis and prediction of land cover dynamics. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 1919–1930. [[CrossRef](#)]
32. Sakamoto, T.; Yokozawa, M.; Toritani, H.; Shibayama, M.; Ishitsuka, N.; Ohno, H. A crop phenology detection method using time-series modis data. *Remote Sens. Environ.* **2005**, *96*, 366–374. [[CrossRef](#)]
33. Gao, F.; Hilker, T.; Zhu, X.L.; Anderson, M.C.; Masek, J.G.; Wang, P.J.; Yang, Y. Fusing landsat and modis data for vegetation monitoring. *IEEE Geosci. Remote Sens. Mag.* **2015**, *3*, 47–60. [[CrossRef](#)]
34. Atzberger, C.; Eilers, P.H.C. A time series for monitoring vegetation activity and phenology at 10-daily time steps covering large parts of south america. *Int. J. Digit. Earth* **2011**, *4*, 365–386. [[CrossRef](#)]
35. Suarez, L.; Zarco-Tejada, P.J.; Gonzalez-Dugo, V.; Berni, J.A.J.; Sagardoy, R.; Morales, F.; Fereres, E. Detecting water stress effects on fruit quality in orchards with time-series pri airborne imagery. *Remote Sens. Environ.* **2010**, *114*, 286–298. [[CrossRef](#)]
36. Pelletier, C.; Webb, G.I.; Petitjean, F. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sens.* **2019**, *11*, 523. [[CrossRef](#)]
37. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 645–657. [[CrossRef](#)]
38. Liang, H.; Li, Q. Hyperspectral imagery classification using sparse representations of convolutional neural network features. *Remote Sens.* **2016**, *8*, 99. [[CrossRef](#)]

39. Li, Y.; Zhang, H.; Shen, Q. Spectral-spatial classification of hyperspectral imagery with 3d convolutional neural network. *Remote Sens.* **2017**, *9*, 67. [[CrossRef](#)]
40. 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](#)]
41. Ji, S.; Zhang, C.; Xu, A.; Shi, Y.; Duan, Y. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sens.* **2018**, *10*, 75. [[CrossRef](#)]
42. Bruce, L.M.; Mathur, A.; Byrd, J.A. Denoising and wavelet-based feature extraction of modis multi-temporal vegetation signatures. *GISci. Remote Sens.* **2006**, *43*, 67–77. [[CrossRef](#)]
43. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: Esa’s optical high-resolution mission for gmes operational services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [[CrossRef](#)]
44. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [[CrossRef](#)]
45. Gitelson, A.A.; Vina, A.; Arkebauer, T.J.; Rundquist, D.C.; Keydan, G.; Leavitt, B. Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophys. Res. Lett.* **2003**, *30*, 4. [[CrossRef](#)]
46. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the modis vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [[CrossRef](#)]
47. Chen, J.; Jonsson, P.; Tamura, M.; Gu, Z.H.; Matsushita, B.; Eklundh, L. A simple method for reconstructing a high-quality ndvi time-series data set based on the savitzky-golay filter. *Remote Sens. Environ.* **2004**, *91*, 332–344. [[CrossRef](#)]
48. Savitzky, A.; Golay, M.J.E. Smoothing+differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [[CrossRef](#)]
49. Rhif, M.; Ben Abbes, A.; Farah, I.R.; Martinez, B.; Sang, Y.F. Wavelet transform application for/in non-stationary time-series analysis: A review. *Appl. Sci.* **2019**, *9*, 1345. [[CrossRef](#)]
50. Mallat, S. *A Wavelet Tour of Signal Processing, The Sparse Way*, 3rd ed.; Academic Press Inc.: Orlando, FL, USA, 2008.
51. Daubechies, I. Ten lectures on wavelets. *Comput. Phys.* **1998**, *6*, 1671.
52. Kaiser, G. A friendly guide to wavelets. *Proc. IEEE* **1995**, *48*, 57–58. [[CrossRef](#)]
53. Lee, G.R.; Gommers, R.; Waselewski, F.; Wohlfahrt, K.; O’Leary, A. Pywavelets: A python package for wavelet analysis. *J. Open Source Softw.* **2019**, *4*, 1237. [[CrossRef](#)]
54. Shannon, C.E. A mathematical theory of communication. *Bell Labs Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
55. Dong, C.; Loy, C.C.; He, K.M.; Tang, X.O. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [[CrossRef](#)] [[PubMed](#)]
56. Masi, G.; Cozzolino, D.; Verdoliva, L.; Scarpa, G. Pansharpening by convolutional neural networks. *Remote Sens.* **2016**, *8*, 594. [[CrossRef](#)]
57. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 142–158. [[CrossRef](#)] [[PubMed](#)]
58. Kamnitsas, K.; Ledig, C.; Newcombe, V.F.J.; Simpson, J.P.; Kane, A.D.; Menon, D.K.; Rueckert, D.; Glocker, B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **2017**, *36*, 61–78. [[CrossRef](#)] [[PubMed](#)]
59. Cheng, G.; Zhou, P.C.; Han, J.W. Learning rotation-invariant convolutional neural networks for object detection in vhr optical remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7405–7415. [[CrossRef](#)]
60. Gong, M.G.; Yang, H.L.; Zhang, P.Z. Feature learning and change feature classification based on deep learning for ternary change detection in SAR images. *ISPRS J. Photogramm. Remote Sens.* **2017**, *129*, 212–225. [[CrossRef](#)]
61. Wang, Q.; Zhang, X.D.; Chen, G.Z.; Dai, F.; Gong, Y.F.; Zhu, K. Change detection based on faster r-cnn for high-resolution remote sensing images. *Remote Sens. Lett.* **2018**, *9*, 923–932. [[CrossRef](#)]
62. Langkvist, M.; Kiselev, A.; Alirezaie, M.; Loutfi, A. Classification and segmentation of satellite orthoimagery using convolutional neural networks. *Remote Sens.* **2016**, *8*, 329. [[CrossRef](#)]
63. Fu, G.; Liu, C.J.; Zhou, R.; Sun, T.; Zhang, Q.J. Classification for high resolution remote sensing imagery using a fully convolutional network. *Remote Sens.* **2017**, *9*, 498. [[CrossRef](#)]

64. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
65. Khatami, R.; Mountrakis, G.; Stehman, S.V. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sens. Environ.* **2016**, *177*, 89–100. [[CrossRef](#)]
66. 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**, *14*, 1685–1689. [[CrossRef](#)]
67. Wang, H.Y.; Zhao, X.; Zhang, X.; Wu, D.H.; Du, X.Z. Long time series land cover classification in china from 1982 to 2015 based on bi-lstm deep learning. *Remote Sens.* **2019**, *11*, 1639. [[CrossRef](#)]
68. 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](#)]
69. He, T.L.; Xie, C.J.; Liu, Q.S.; Guan, S.Y.; Liu, G.H. Evaluation and comparison of random forest and a-lstm networks for large-scale winter wheat identification. *Remote Sens.* **2019**, *11*, 1665. [[CrossRef](#)]
70. Mou, L.C.; Bruzzone, L.; Zhu, X.X. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 924–935. [[CrossRef](#)]



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).