Research Article

Vegetation Coverage Monitoring Model Design Based on Deep Learning

Lilei Zhang, Wenwen Wang, Yujie Dong, and Longliang Wu

1 College of Architecture and Survey and Draw Engineering, Shanxi Datong University, Datong 037000, China
2 Datong Key Laboratory of Mine Geohazard Prevention and Environment Recovery Shanxi, Datong 037000, China
3 Key Laboratory of Coal Geological Geophysical Exploration Surveying & Mapping Institute of Shanxi Province, Jinzhong 030600, Shanxi, China
4 bureau Public Works of Shenzhen Municipality, Shenzhen 518031, Guangdong, China

Correspondence should be addressed to Longliang Wu; wulongliang@foxmail.com

Received 26 May 2022; Revised 9 June 2022; Accepted 13 June 2022; Published 21 July 2022

Academic Editor: Lianhui Li

Based on the medium-resolution Landsat TM and OLI satellite images in the study area, the deep learning ENVINet-5 model is adopted for vegetation coverage monitoring. By referring to the fusion image and Google Earth high-resolution satellite image, the training samples and verification samples are manually labeled, and the labels of four types of ground objects (desert, water body, cultivated land, and construction land) are made. Through the ENVI deep learning binary classification model, the labeled training samples are trained, and a large number of samples of desert, water, and cultivated land are extracted and transformed into corresponding label images. Then, a large number of training sample labels extracted from the model are combined with the manually made construction land sample labels and both of them are used as the training samples of the ENVI deep learning multiclassification model. According to the classification process of the deep learning model (creating label image, initializing training model, and training model and model classification), through the adjustment of various parameters, the four types of ground objects in the study area are finally classified. Finally, the classification results that meet the accuracy requirements are statistically analyzed. It is proved that the model classification results can meet the use requirements.

1. Introduction

Desertification is an important factor threatening human survival and sustainable development in the 21st century. How to effectively prevent and mitigate desertification is the focus and hotspot of current research [1–4]. Land desertification seriously reduces the carrying capacity of the ecological environment, and the area and degree of desertification land are rapidly expanding, which seriously restricts the sustainable development of society [5]. Monitoring vegetation coverage and rocky desertification and timely and accurate assessment of the current situation and changes in desertification will help to formulate global actions to prevent and eliminate desertification [6–8].

The traditional desertification research mainly adopts the artificial field mapping method, which has the disadvantages of time-consuming, laborious, and low efficiency in the implementation process [7]. With the emergence of remote sensing technology [9–11] and vigorous development of information technology [12–18], its macro-, objective, and economic advantages make up for the shortcomings of traditional desertification research to a certain extent, so it can better deal with environmental change monitoring [16–18]. On the other hand, with the development of computer technology, the machine learning method [19–21] is widely used in the classification and extraction of remote sensing images. It can better extract the ground features in the image. Especially in recent years, the emergence of deep learning has greatly improved the technical means of image processing. Rapid and accurate feature classification of a large amount of data can better adapt to the development of society and meet the needs of human production and life.
As a complex form of machine learning, deep learning can make the system automatically discover the feature expression of data. Compared with other machine learning types, it can continuously improve the prediction accuracy without external guidance or intervention, and draw conclusions through multilayer learning in neural networks. For the processing of remote sensing images, deep learning attempts to discover and use the spatial, spectral, and statistical features of remote sensing images. In view of this, a vegetation coverage and rocky desertification monitoring method is proposed based on deep learning in this study, so as to provide a scientific basis and data basis for ecological environment monitoring and sandy vegetation restoration in desertification areas.

2. Methods

2.1. Deep Learning Model Architecture. The ENVINet-5 model is developed based on TensorFlow deep learning framework. The model architecture is based on the improvement of the U-NET neural network, which is similar to the U-NET architecture. It is a mask-based and encoder-decoder architecture. Combined with the powerful remote sensing image processing software ENVI, researchers can directly use the deep learning network to process remote sensing images.

This study uses the ENVINet-5 deep learning module to classify the ground objects in Landsat medium-resolution satellite images in the study area [22, 23]. The ENVINet-5 model architecture has 5 levels and 27 convolution layers, and each level represents different pixel resolutions in the model. We input the original image into the model, slice it, and then use a \( 3 \times 3 \) convolution layer to convolute the slice, so as to increase the number of image features. We use a \( 2 \times 2 \) pooling layer to reduce the size of the image, so as to reduce the operation and retain most of the features of the image. The sampling process is divided into upsampling process and downsampling process. We recognize and classify many features generated in the downsampling process, so as to achieve the purpose of feature classification.

After the remote sensing image is loaded into the deep learning network, the convolution operation will be carried out after slicing. Convolution is essentially to extract the features in the image and generate feature images with different dimensions, which are often composed of multiple convolution kernels of different sizes. The size of the commonly used convolution kernel is \( 3 \times 3 \) and \( 5 \times 5 \). In this study, the convolution kernel size of a deep learning network is \( 3 \times 3 \). In addition, the function of activation function is indispensable in convolution operation. The activation function can better solve the gradient disappearance problem, simplify the computational complexity, speed up the training speed, and reduce the burden of computer operation. In convolution operation, the rule activation function is often used. As a nonlinear activation function, it can better solve the problem of overfitting in training.

After the convolution operation, a large number of feature images will be generated. If the convolution operation is carried out again, the amount of operation will be greatly increased. Therefore, it is necessary to reduce the amount of computation and retain more feature information of the generated feature image. The appearance of the pooling layer solves this problem better. Common pooling methods include mean pooling and max pooling. In this study, max pooling is used, and the window size is \( 2 \times 2 \). It can not only extract the main features but also reduce the amount of computation. After pooling, the feature image size is reduced to half of the original size.

In addition, the loss function is introduced into the deep learning network, which can map values to non-negative space. When the value of the loss function is smaller, it means that it is closer to the real value. Therefore, the classification effect of the model can be quantitatively evaluated and optimized. The commonly used loss functions include root mean square error, cross-entropy loss, and so on.

2.2. Deep Learning Model Parameters

2.2.1. Epochs and Batches. In order to make the training model achieve better classification results, the number of training iterations (epochs) needs to be adjusted. In ENVI deep learning, the number of iterations (epochs) represents that the model intelligently extracts the number of slices (batches) from the label image. Because there is bias judgment when extracting slices, an epoch refers to the number of slices trained before bias judgment adjustment.

The determination of the number of epochs and the number of extracted slices has certain randomness, which depends on the diversity of learning feature sets. In ENVI deep learning, the number of iterations is generally set between 16 and 32. For the number of patches per epoch, the number of slices extracted by each epoch determines the training amount of the model, which is usually set between 200 and 1000.

The model will use multiple slices at the same time in an iterative training. Batch refers to a group of slices read in an iterative training. Batches run in an iterative training and stop after reaching or exceeding the total number of slices. The number of slices in a group needs to be set according to the size of GPU video memory.

2.2.2. Class Weight and Loss Weight. The ENVI deep learning network uses the statistical technology of inverse transform sampling to train the model based on slice deviation selection. By introducing a deviation, the model can see the slices of highlighted feature pixels more often. In inverse transform sampling, the samples introduced into the model are directly proportional to their contribution to the probability density function. The deviation is controlled by the class weight parameter in the Train TensorFlow Mask Model tool. The value range of the class weight parameter is 0 to 2 (when the maximum value is set for the sparse training set, the valid range of the maximum value is 0 to 6).

The loss weight parameter is used for the deviation loss function, focusing on the correct identification of characteristic pixels and reducing the identification of background
pixels. Parameters are useful when the distribution of feature targets is sparse. A value of 0 indicates that the model treats the feature and background pixels equally. Increasing the value of the loss weight parameter will bias the loss function to find the characteristic pixels. The valid range of parameter values is 0 to 3.

2.2.3. Solid Distance and Blur Distance. In the process of model training, in addition to the weight of features and background pixels, the size and edge of features must also be considered. The solid distance parameter is often used to expand the size of linear or point features. The parameter value size represents the number of pixels around the marked ROI, which forms the target feature together with the ROI.

In deep learning, the setting of blur distance helps the model to obtain the sharp edge of the target. By setting the fuzzy edge and gradually reducing the fuzzy value in the model training, the model is close to the sharp feature of the target. In the ENVINet-5 model, the value range of blur distance is 0 to 70.

2.3. Deep Learning Model Extraction Process. The ENVINet-5 is defined by basic neural network parameters. To extract target, it is necessary to create a label raster that can indicate the target and then use the label sample to train the model. The label image can be created jointly with the tools ROIs and Deep Learning/Build Label Raster from ROI, and then, the label sample is used to train the model. The trained model can find targets with similar features in other images. The extraction result is a class activation map/raster, in which the DN value represents the probability that the pixel belongs to a certain ground object target. The deep learning training model involves the adjustment of parameters and has a certain degree of randomness. Due to the convergence mode of the algorithm, different models will be obtained even if the parameters in the training are the same. The extraction process is shown in Figure 1.

2.3.1. Create Label Image. For target recognition, label rasters need to be created to train the model. Based on the original image to be classified, through field investigation, sample site selection or, combined with a high-resolution image as reference, we establish proprietary identification marks for different ground objects, so as to create label images. In order to generate high-quality label images, representative samples of typical areas should be selected.

First, we select the object of interest in the typical area to create ROI. The ROI drawn should contain a variety of surface features such as shape, color, and texture, which is conducive to improving the classification accuracy of the model.

Finally, we use the Deep Learning/Build Label Raster from ROI tool in the ENVI toolbox to create label images. The tag image contains the original band and mask band of the input image. Different DN values of the mask band represent different types of ground objects.

2.3.2. Initialize TensorFlow Model. In ENVI, we select the Deep Learning/Train TensorFlow Mask Model tool. Before deep learning model training, it is needed to initialize or load a TensorFlow model. Model initialization needs to define model parameters, including patch size, number of bands used for training, and saving path.

2.3.3. Train TensorFlow Model. After model initialization, various parameters and weights need to be set in the train TensorFlow mask model to guide TensorFlow model to learn target features.

In the training process, the label image is repeatedly exposed to the model. The model learning converts the spectral and spatial information in the label image into a class activation gray image to highlight the extracted target. Through the loss function, the model can know the wrong random guess results. Through the adjustment of internal parameters or weights of the model, the model is more accurate. The trained TensorFlow model can find the same features in other images.

2.3.4. Execute Classification. We use the trained model to search for the same ground feature in other images and use TensorFlow Mask Classification tool to classify and extract remote sensing images.

2.4. Remote Sensing Monitoring of Vegetation Coverage. Vegetation coverage refers to the percentage of the vertical projection area of vegetation on the ground in the total area of the statistical area. It is of great practical significance to analyze and evaluate the regional ecological environment by obtaining the regional surface vegetation coverage and revealing the change and dynamic change trend of surface vegetation in the region. The vegetation index method has little dependence on surface measured data, has universal applicability, and can be extended to a wide range of areas. It is a reliable means to quantitatively monitor the change in vegetation cover.

The calculation formula for normalized vegetation index is as follows:
\[
NDVI = \frac{(NIR - RED)}{(NIR + RED)}
\]

where \(NIR\) is the near-infrared reflectance of remote sensing image, and \(RED\) is the infrared reflectance of remote sensing image.

Vegetation coverage retrieved by normalized vegetation index is as follows:

\[
NDVI_s = \frac{(VFC_{max} \times NDV_{min} - VFC_{min} \times NDV_{max})}{(VFC_{max} - VFC_{min})}
\]

\[
NDVI_v = \frac{(1 - VFC_{min}) \times NDV_{max} - (1 - VFC_{max}) \times NDV_{min})}{(VFC_{max} - VFC_{min})}
\]

where \(NDV_{max}\) and \(NDV_{min}\) are the maximum and minimum NDVI values in the study area. Because noise is inevitable, the value within a certain confidence interval is generally taken, and the confidence interval is determined according to the actual situation of the image.

Furthermore, it can be converted into the following:

\[
VFC = \frac{(NDV_I - NDV_{Imin})}{(NDV_{Imax} - NDV_{Imin})} \tag{4}
\]

Therefore, the data of different vegetation coverage levels in each period of the study area can be obtained.

2.5. Deep Learning Model. Based on the U-shaped neural network structure, we use the means of jump connection to make the output UAV image features of each decoding layer in the decoder corresponding to the encoder (the image size is the same). Each decoding layer is an interworking structure, which makes full use of the image features extracted by all coding layers in the encoder. Compared with the traditional neural network deep learning model based on encoder-decoder, it can get more accurate results under a small amount of training data. The specific model framework is shown in Figure 2.

As can be seen from Figure 2, the encoder part of the model is designed with four coding layers and two convolution layers, and the decoder is composed of four decoding layers and one softmax layer, in which the activation function used by the model is ReLU function. The specific steps are as follows: input the remote sensing HD image data into the encoder, extract the features through the encoder, and output the feature images with the size of 1/2, 1/4, 1/8, and 1/16 (1 represents the image size), with the numbers of 16, 32, 64, 128, and 256, respectively. Then, the feature images output by the encoder are analyzed by the decoder and the image segmentation results with the size of 1 are output, that is, the divided vegetation coverage area (the result output is white) and nonvegetation coverage area (the result output is black), so as to calculate the vegetation coverage of different typical sample plots.

3. Experiment and Analysis

This research mainly applies the computer to test. The computer system is Windows 1064 bit professional version, the processor model is Intel Core i7-9750h, the graphics card model is NVIDIA GeForce GTX 1050, the memory size of the solid-state disk is 512 g, and the memory size is 8 g. The computer is loaded with the remote sensing image processing software ENVI 5.5, in which the deep learning binary classification model and multiclassification model are installed for the binary classification and multiclassification of ground objects, respectively.

In the process of training model, the label image is needed to mark the training target so that the ground object categories can be distinguished in the neural network. We use the region of interest tool in ENVI to manually draw or import the existing vector data. At present, the deep learning two classification model supports single-class target extraction. In view of the need for a large number of training samples for deep learning, and according to the actual distribution of features in the study area, label images are made for deserts, water bodies, and cultivated land in the study area. After model extraction, a large number of samples can be provided for the deep learning multiclassification model.

The study selects the deep learning binary classification model in ENVI and uses ENVINet-5 to extract desert, water, and cultivated land from some label samples. Due to the small area occupied by the construction land, it is manually selected with reference to the high-score image. Then, the extracted three types of features and construction land are made into multiclassification label samples as the training and verification samples of the deep learning multiclassification model. Figure 3 shows a partial screenshot of the desert, water, and cultivated land label images.

After the image label is made, the model needs to be initialized. We need to define model parameters, including patch size, the number of bands used in training, and output path. The slice size of this test is set to 300 x 300 pixels, which cannot be greater than the number of pixels with the minimum side length of the cropped subregion, and the number of bands is set to the number of bands of the original.
image. After the initialization of the model, it is necessary to set the parameters of the training model (train TensorFlow mask model). For the number of iterations, we set the number of iterations to 30 according to the computing power of the computer processor and the size of GPU memory. The number of slices per epoch determines the training amount of the model. The number of slices is generally selected between 200 and 1000. We set the number of slices in iterative training to 300. The number of patches per batch is set to 4. The fixed distance and blur distance are mainly set according to the drawing of the label image. The classification weight has a minimum value of 0 and a maximum value of 2, and the loss weight ranges from 0 to 3. We adjust it according to the actual training effect.

After parameter adjustment, the final parameter is set as 20 iterations, 300 training slices per iteration, 4 reading slices per batch of training, 0.8 classification weight, and 0.6 loss weight. Considering the features of the study area, we do not set the fixed distance and fuzzy distance. The image extraction results are shown in Figure 4.

The desert, water body, and cultivated land extracted by the binary classification model are used to make the region

**Figure 2:** The deep learning model for vegetation coverage extraction.

**Figure 3:** A partial screenshot of the desert, water, and cultivated land label images.
of interest based on the extraction results. Together with the manually drawn construction land, the area of interest is shown in Figure 5.

The training process of the multiclassification model is similar to that of the two classification models, with occasional differences. Only when initializing the output category of the training model, we can set it to 4 categories and set the number of slices to 304. The multiclassification depth model trained by samples is used to extract different features in the study area. The loss function curve and training accuracy curve in the training process are shown in Figures 6 and 7. The minimum loss value is 0.1483, and the maximum training accuracy is 0.9425.

After extraction, the classification map of desert, water body, cultivated land, and construction land is obtained, and then, the accuracy test is carried out in combination with the verification points selected by field investigation, fusion image, and Google Earth high-score image. A total of 43646 validation samples (20055 deserts, 9169 water bodies, 9407 cultivated land, and 5015 construction land) were used to calculate the producer accuracy, overall classification accuracy, and kappa coefficient of feature classification in combination with the transfer matrix to carry out the comprehensive evaluation model. The overall classification accuracy is equal to the sum of correctly classified pixels divided by the total number of pixels. Kappa coefficient represents the proportion of error reduction in the evaluated classification compared with completely random classification. We output the classification results that meet the precision requirements. The test results are shown in Table 1.

Based on the verification accuracy of the extraction results, we can see that the overall classification accuracy is
more than 85%, and the kappa coefficient is about 0.85. The model classification results can meet the use requirements.

4. Conclusions

In this study, we take the northwest region of China as the research area. Based on the Landsat medium-resolution remote sensing satellite images of five periods (2000, 2005, 2010, 2015, and 2018) in the study area, and based on ENVI 5.5, the remote sensing image software processing platform used for image preprocessing, the deep learning model is used to classify the features of desert, water body, construction land, and cultivated land in the study area in recent 19 years. The classification results satisfying the accuracy are obtained; furthermore, the annual change in desertification in the region is statistically analyzed by using the single land dynamic degree and land-use transfer matrix method, and the mutual transfer between desert and other features. Based on the above classification results, the desert area was extracted to obtain the normalized vegetation index (NDVI), and then, the vegetation coverage of the study area was obtained. Combined with the quantitative relationship between vegetation coverage and desertification, the classification results of desertification grades in different years in the study area were obtained, and the mutual transformation among severe, moderate, and mild desertification areas was analyzed by using the transfer matrix method. The results show that the main feature of the study area is the rocky desert, accounting for more than 80% of the area. In the past 19 years, it has reduced to 106.33 km², the average annual growth rate is −0.32%, and the rocky desertification continues to reverse. The results showed that the ecological environment in different research periods developed well as a whole.

Compared with traditional machine learning classification methods, the deep learning method used in this study can automatically extract the features of remote sensing images, so it can effectively avoid the complex work of manual extraction and feature selection, and has high operability.

There are many influencing factors of desertification. Due to the limitations of data source, time, economy, and technology, this study still has many deficiencies. The Landsat series remote sensing data used in this study do not have a high spatial resolution, and one pixel may contain a variety of ground objects. It is not conducive to the establishment of a pure feature sample set, which has certain errors in the training and verification of the model. The high spatial resolution remote sensing data greatly increase the probability of pure pixels, which is conducive to the better learning of samples by the deep learning model. In the model training of deep learning, the adjustment of parameters is subjective and tentative. After meeting the classification results, further experimental research was stopped. In the next step, the scientific parameter adjustment method can be used to obtain better classification results.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.
Acknowledgments

This work was financially supported by the College-Level Project of Shanxi Datong University (XJG2021223) and the Teaching Reform Project of Shanxi Datong University (2020K14).

References


