DESERTIFICATION GLASSLAND CLASSIFICATION AND THREE-DIMENSIONAL CONVOLUTION NEURAL NETWORK MODEL FOR IDENTIFYING DESERT GRASSLAND LANDFORMS WITH UNMANNED AERIAL VEHICLE HYPERSPECTRAL REMOTE SENSING IMAGES

W. Pi, J. Du *, H. Liu, X. Zhu

Inner Mongolia Agricultural University, Mechanical and Electrical Engineering College, Hohhot, China; e-mail: nndjwc202@imau.edu.cn

Based on deep learning, a desertification grassland classification (DGC) and three-dimensional convolution neural network (3D-CNN) model is established. The F-norm^2 paradigm is used to reduce the data; the data volume was effectively reduced while ensuring the integrity of the spatial information. Through structure and parameter optimization, the accuracy of the model is further improved by 9.8%, with an overall recognition accuracy of the optimized model greater than 96.16%. Accordingly, high-precision classification of desert grassland features is achieved, informing continued grassland remote sensing research.

Keywords: *unmanned aerial vehicle, hyperspectral image, grassland desertification, ground object identification, 3D convolutional neural networks.*

МОДЕЛЬ ТРЕХМЕРНОЙ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ ИДЕНТИФИКАЦИИ РЕЛЬЕФА ПУСТЫННЫХ ПАСТБИЩ С ПОМОЩЬЮ ГИПЕРСПЕКТРАЛЬНЫХ ИЗОБРАЖЕНИЙ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ

W. Pi, J. Du *, H. Liu, X. Zhu

УДК 551.507.321.2

Сельскохозяйственный университет Внутренней Монголии, Хух-хото, Китай; e-mail: nndjwc202@imau.edu.cn

(Поступила 25 июня 2019)

Создана гиперспектральная система дистанционного зондирования с беспилотного летательного аппарата для исследования пустынных пастбищ Внутренней Монголии (Китай) при естественном освещении в полевых условиях. На основе машинного обучения предложена модель трехмерной сверточной нейронной сети (3D-CNN) для классификации пустынных пастбищ. Для уменьшения объема данных использована парадигма F-norm^2 при обеспечении целостности пространственной информации. Благодаря оптимизации структуры и параметров модели ее точность дополнительно повышается на 9.8%, при этом общая точность распознавания оптимизированной модели >96.16%. Соответственно достигается высокоточная классификация признаков пустынных пастбищ, что способствует повышению эффективности исследований по дистанционному зондированию пастбищ.

Ключевые слова: *беспилотный летательный аппарат, гиперспектральное изображение, опустынивание пастбищ, идентификация наземных объектов, трехмерная сверточная нейронная сеть.*

Introduction. China's natural grassland area is 390 million hm^2. It accounts for 40% of the country's land area, which is more than cultivated land and forest areas combined. Furthermore, this accounts for 12% of the global grassland areas, ranking as the largest national grassland area in the world. However, 80% of China's grasslands are distributed in northern China [1]. The Inner Mongolia grassland is a notable traditional animal husbandry base and green ecological barrier in northern China. Because of long-term use and global climate change, the Inner Mongolia grassland has encountered severe ecological challenges during the past half century [2]; productivity and ecological functions are significantly reduced. The area of desertifica-

tion in Inner Mongolia is greater than 60 million hm^2, accounting for approximately one-fourth of the national desertification area, of which 98% is comprised of grassland areas [3]. The area of desertification distribution is wide and the degree of degeneration is different. Vegetation coverage is a functional indicator for visual and quantitative estimations of surface plant communities; used to describe vegetation dynamics, it indicates regional ecological environment changes [4]. Monitoring changes in vegetation coverage has become an important means to evaluating regional ecological environmental quality. Traditionally, vegetation coverage is monitored through manual methods, including visual estimation methods and instrument methods [5]. Not only do such measurement methods have many limitations, but they also require significant time and material resources [6]. With the development of optical, thermal infrared, and microwave technologies, the use of satellite-borne sensors to observe ground objects, including the multispectral scanner system (MSS), satellite for observation of earth (SPOT), and thematic mapper (TM), has gradually increased. Remote sensing data can achieve a wide range of measurements for vegetation coverage [7–9]. However, because of the distance between satellite sensors and information on the ground, the obtained vegetation information is limited by the spatial resolution of the image. Thus, the identification and inversion of dense vegetation coverage areas can be obtained; however, with satellites alone, it is difficult to achieve fine identification and high-precision in vegetation information in desert grassland. To obtain more precise vegetation information, optical sensors have been increasingly combined with drones. This forms a low-altitude drone remote sensing platform. Such unmanned aerial vehicle (UAV) remote sensing platforms not only fulfill the gap in information between measurements on the ground and satellite remote sensing image, but also ensure data timeliness [10, 11]. Accordingly, low-altitude hyperspectral remote sensing satisfies the requirements of data accuracy for desertification grassland while also allowing flexibility and efficiency in data collection.

Since deep learning was introduced by Hinton et al. in 2006, deep learning has been successfully applied to machine learning, including its excellent performance in recognition [12], medical image recognition [13], and image classification [14]. With the success of deep learning in natural image processing, deep learning methods have been gradually introduced into remote sensing image classification [15]. Remote sensing image classification is characterized by three types of classification: spectral feature classification, spatial feature classification, and spectral-spatial feature classification [16, 17]. Among these, spectral feature classification developed first, as the spatial dimensions of early sensors were very limited [18]. With further sensor development, spatial dimension information greatly increased. The classification methods of spectral features then gradually developed into the joint use of spectral-spatial information. It has been shown that spatial context information helps to explain the relationship between adjacent pixels; utilizing the entirety of spatial information effectively improves classification accuracy [19, 20]. Scholars have since used a combination of spectral and spatial properties in many studies, one of which involves the initial separate extraction of spectral and spatial features, prior to the combination of the spectral and spatial features [21]. Separately extracting and then combining spectral and spatial features is often performed with a stacked autoencoder (SAE), deep brief network (DBN), one dimensional convolution neural network (1D-CNN), or two-dimensional convolution neural network (2D-CNN). In another method, which directly involves hyperspectral imaging (HSI), the depth spectral spatial features are extracted from the 3D cube, and a three-dimensional convolution neural network (3D-CNN) is used to process image cubes [22–24]. At present, the SAE, DBN, and CNN models are used to classify and identify HSI data captured by various sensors [25]. The classification accuracy of the 3D-CNN model is significantly better than the SAE and DBN models. SAE, DBN, and other models applied earlier in HSI image classification efforts have since significantly improved, allowing HSI images to achieve sufficient results [22, 24]. However, with high-precision recognition of objects and the study of spectral-spatial characteristics with higher precision requirements, the 3D-CNN shows great potential.

Currently, most investigations have used satellite and aerial remote sensing image data to concentrate on the identification and classification of large-scale crops, buildings, roads, and forests. These are limited by satellite and aerial sensor accuracy, and the image spatial resolution is low. It is difficult to achieve high-precision identification and inversion of small features; while the desert grassland vegetation is sparse, the coverage is less than 40%, the average height is less than 8 cm, and the distribution is random and scattered. Accordingly, a high-precision method for data collection with a deep learning model is proposed. This study constructs a low-altitude UAV remote sensing platform to collect remote sensing images of desertification grasslands. This ensures data timeliness and improves data accuracy. A 3D-CNN is used to construct the desertification steppe vegetation coverage model. Data structure dimension reduction, model structure, and parameter tuning are used to obtain the best grassland land object recognition model. In the development of

this high-resolution remote sensing UAV platform for the desertification grassland data collection, innovative methods for grassland remote sensing research are developed.

This study contributes the following:

1. A UAV hyperspectral remote sensing platform is constructed to collect data for desertification grasslands.

2. Based on 3D convolution, a desertification steppe General Direction of Development Cooperation (DGC) 3D-CNN model is established for low-altitude hyperspectral remote sensing data.

3. F-norm^2 is used for data dimension reduction. By comparing the overall accuracy and training duration of the five dimensions of data, the best dimension for adapting the model is determined.

4. The structure and parameters of the DGC-3D-CNN model are optimized, the network performance effectively improved, and a comparably superior desertification grassland classification model obtained.

Materials and methods. *Overview of the study area.* The experimental area is located in the Gegentala Grassland (41°78′43″N, 111°87′41″E) in the middle and western area of Siziwang Banner, Wulanchabu City, Inner Mongolia, China. This area is a typical desertified grassland with a temperate continental climate: the average annual precipitation is 280 mm, 75% of which occurs from June–August; the annual average temperature is 3.4°C; the soil type is mainly pale chestnut soil, and the group is short. From *Stipa breviflora*, the dominant species are *Artemisia frigida* and *Cleistogenes mutica*; the main associated species are *Convolvulus ammannii*, *Neopal-lasia pectinata*, and *Allium tanuissmum*. Vegetation grass layers are low, sparse, and random; they have an average height of less than 8 cm and a coverage rate of only 13–35% [26, 27].

Construction of UAV Hyperspectral Remote Sensing System. The UAV hyperspectral platform is primarily composed of an eight-rotor UAV, a hyperspectral imager, a pan-tilt, and an inertial navigation system. The hyperspectral instrument is a Pika XC2 hyperspectral instrument manufactured by Resonon, USA. The HEX-8 UAV is produced by Jinan Share UAV in China and adopts the professional A3pro flight control system, which can achieve autonomous flight of the route. The maximum takeoff weight is 40 kg, and the full flight duration is 25 min; the Head is adopted from SZ DJI Technology Co., Ltd. RONIN-M Head, manufactured by SZ DJI Technology Co., Ltd., has a built-in DJI advanced 32-bit DSP processor and weighs 3.6 kg. The inertial navigation system is the Ellipse miniature inertial sensor produced by SBG of France. The main parameters are follows: lens focal length 17 mm, field of view 30.8°, spectral range 400–1000 nm, spectral resolution 1.3 nm, number of spectral channels 447, number of spatial channels 1600, pixel size 5.86 μm, scanning method – linear push sweep, maximum flight speed 10 m/s, flight mode – route autonomous flight, weight 32 kg, endurance (time) 25 min.

Data collection. According to the climate characteristics of the grassland and the growth cycle characteristics of the grass, the data collection period is from August 20–August 25, 2018. The weather conditions involved a wind of less than 3 level. Between 10:00 and 14:00, hyperspectral data was acquired by flight. To ensure that the data are not overexposed or underexposed because of changes in the cloud amount, standard reference whiteboard calibration is performed before and after each flight. The UAV-HSI system uses autonomous flight along the route to collect remote sensing images for 2.5 hm^2. To ensure image quality and data accuracy, the single image size is set to 3640 lines×1600 samples×231 bands, and the flying height of the drone is 30 m. The flight speed is 1 m/s, the side overlap is 55%, and the image spatial resolution is approximately 2.1 cm. According to the calculation, it takes 3 trips to collect 2.5 hm^2, wherein the single flight times are 20, 20, and 12 min. To ensure data quality, the test area is collected 3 times, for a total of 9 flight times.

Data preprocessing. First, the best image quality was selected by manual inspection, and the radiance correction was performed using Spectral Pro software. After data correction, the image includes 400–1000 nm, a total of 231 bands, a spatial resolution of 2.1 cm/px, and a size of 1600×3603 px. The corrected image is shown in Fig. 1a. The image is divided into three types of ground truth classes (Fig. 1b). Of the labeled samples, 40% were selected as training data, and the rest were used for testing (Table 1).

TABLE 1. Land-Cover Classes and Numbers of Samples in the Part-UAV-HSI

No.	Class	Total	Test	Training
1	Vegetation	3,279,270	1,967,562	1,311,708
2	Soil	1,620,862	972,517	648,345
3	Others	2,268	1,360	908
All classes		4,902,400	2,451,200	2,451,200

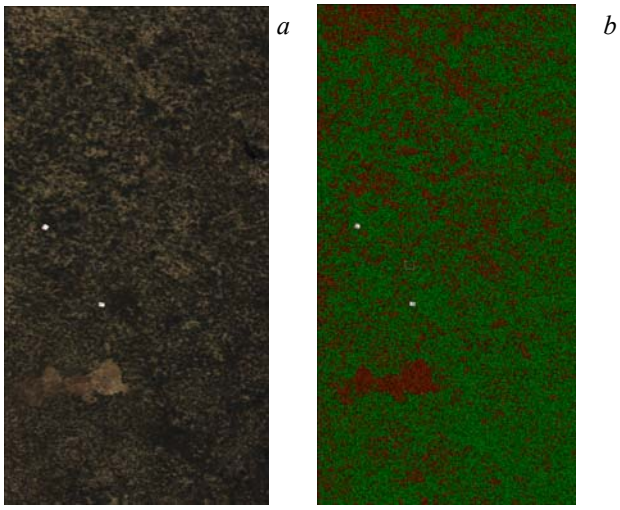

Fig. 1. (a) UAV-HSI and (b) ground truth.

DGC-3D-CNN model and F-norm^2 band selection. *3D-CNN.* With the integral convolutional layer of the CNN structure, convolution operations are often used to extract features and introduce nonlinear factors into the network through activation functions. The convolution kernel of a 3D-CNN is a three-dimensional cube. For hyperspectral data, the convolution kernel slides on the band while sensing the entire channel of the field of view image, thus achieving the calculation of spectral and spatial components. When there are fewer training parameters, the information provided in the data is utilized as much as possible, and the cost is fundamentally reduced. The 3D convolution is expressed as

$$\upsilon_{kmn}^{\delta\varepsilon\epsilon} = f\left(\sum_{l}\sum_{l=0}^{L-1}\sum_{w=0}^{W-1}\sum_{h=0}^{H-1} u_{kn}^{hwd} v_{(n-1)m}^{(\delta+h)(\varepsilon+w)(\epsilon+h)} + q_{kn}\right), \quad (1)$$

where l, w, and h represent the length, width, and height of the convolution kernel cube, respectively, and m and n respectively represent the kernel of the previous layer and the current layer, respectively; $\upsilon_{kmn}^{\delta\varepsilon\epsilon}$ is the $(\delta,\varepsilon,\epsilon)$th value calculated by convolving the mth feature cube of the preceding layer with the nth kernel of the kth layer, and u_{kn}^{hwd} is the previous position, (h, w, d), calculated by convolving the mth feature cube in the preceding layer [28].

DGC-3D-CNN model. The Python language is used to construct the DGC-3D-CNN network (Fig. 2) using the deep learning PyTorch framework. According to the empirical value and initial value obtained by the research, the initial network framework is composed of two 3D convolution layers, two pooling layers, and one fully connected layer. The activation function is ReLU. The main parameters are shown in Table 2, where 3*3*3@20 represents 20 three-dimensional convolution kernels with a size of 3*3*3, a batch size of 256, a patch size of 9, and a full-link unit number of 1024.

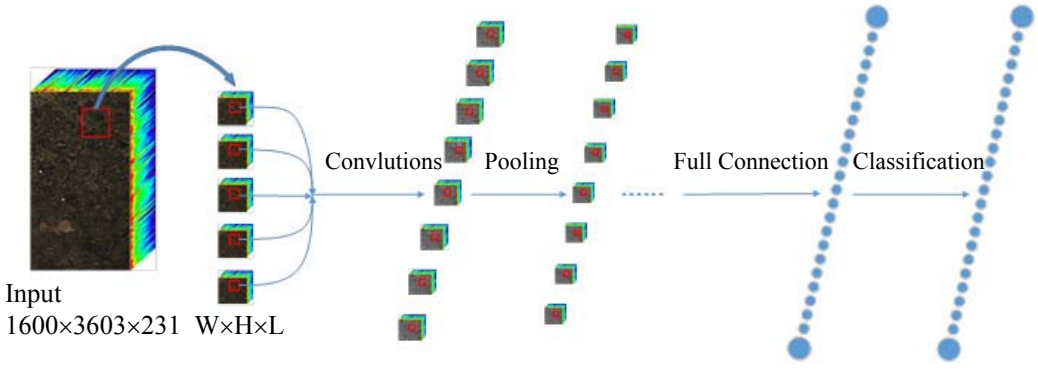

Fig. 2. Schematic diagram of DGC-3D-CNN network structure.

TABLE 2. Parameters in the DGC-3D-CNN Method

Network layer	Convolutional layer	Act-F	Pooling layer	Pooling function
1	3*3*3@20	ReLU	3*1*1	Max-pooling
2	3*3*3@35	ReLU	3*1*1	Max-pooling

Data dimension reduction. HSI uses a large number of bands and narrow band spacing, which can provide rich spectral characteristics and spatial characteristics. This provides high-precision ground object recognition, but also complicates and lengthens the data learning process. The data collected by the UAV remote sensing platform is 1600*3603, with a total of 231 dimensions and a band spacing of 4.32 nm. The band selection should reduce the data dimension as much as possible while eliminating useless information and noise interference, but also retain the complete spectral space information as much as possible. Based on this, the Frobenius norm (F-norm), is used [29]

$$X(:,:,b)_F = \left(\sum_r^R \sum_c^C \left| X(r,c,b) \right|^2 \right)^{\frac{1}{2}}, \tag{2}$$

where X is a tensor, R is the number of rows of tensor, C is the number of columns of tensor, and b is the dimension of the tensor. The values R is 1, 2, ..., 1600. The values C is 1, 2, ..., 3603, and the values B is 1, 2, ..., 231. The F-norm value corresponding to the image represents the energy of the corresponding band. If the F-norm^2 value is too small, the amount of information contained therein is too low. If the F-norm^2 value is too large, the noise is severe. Derived from Eq. (2), the square of the F-norm values of the respective bands in the UAV-HSI image are shown in Fig. 3. As can be seen, the amount of information contained in the first eight bands is low. The F-norm^2 value in bands 198–231 band sharply increases because of noise, which should be eliminated. Accordingly, the remaining 9–197 bands are selected and recorded as id1. Taking the F-norm median band, the 104th band, as the central band, the two sets of bands are selected in increments of 20 bands: bands 86–126, 66–146, 46–166, and 26–186, and the remaining bands are respectively recorded as id2, id3, id4, id5, and id6. These groupings of band data are used as input data to classify the overall accuracy (OA), the kappa value, and the model training time of the DELL-T7920 workstation as an evaluation index. Additionally, different band groupings are compared, and the price and performance of the quantity of input data evaluated.

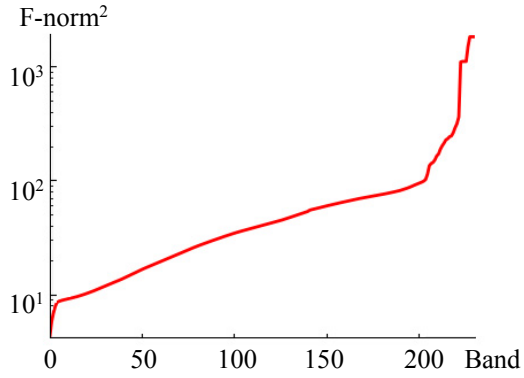

Fig. 3. Square of the Frobenius norm of UAV-HSI.

TABLE 3. Calculation Effects of Different Band Models

No	Bands	Total bands	Kappa	OA, %	*T*, min	DV, %	T-DVR, %
id1	9-197	188	0.805	88.12	378	–0.97	+7
id2	86-126	41	0.816	86.35	270	–2.68	–23
id3	66-146	81	0.827	87.62	305	–1.47	–13
id4	46-166	121	0.834	**88.97**	**326**	**–0.06**	**–7**
id5	26-186	161	0.837	**89.03**	**351**	–	–
id6	1-231	231	0.731	82.32	462	–6.71	+31

The results of the model calculation are shown in Table 3. Therein, the id6-full-band data classification requires the longest amount of time and the classification effect is the worst; the id1 classification accuracy is higher than id6, but it requires the longest amount of time; as the number of bands increases, the id2–id5 kappa value, the OA, and the time consumed all increase. The best performance is achieved by id5 with an OA of 89.03%, and a time of 351 min. To compare the accuracy and time consumption, id5 is used as the reference value to obtain the time difference as well as the precision OA difference (OADV), such that OADV= idnOA-id5OA. The ratio of time to OADV (T-DVR), such that T-DVR=(idnT-id5T)/id5T, is shown in Table 4. Accordingly, the best performing id5 is only 0.06% higher than the id4 precision value, but the time consumption is increased by 7%. The next id4 has a higher price/performance ratio.

DGC-3D-CNN structural parameter optimization. The DGC-3D-CNN model is based on empirical values and initial parameters. The pseudo code for the model is shown in Table 4. The performance of the CNN model is affected by the parameter settings; if the 3D-CNN parameter settings are not appropriate, the network performance will be greatly reduced [25]. To optimize the network performance, id4 is used as the input data set, and the nine parameters in DGC-3D-CNN are adjusted according to the logical relationship of parameters and the univariate principle. The optimal parameters are selected according to the OA of the classification.

TABLE 4. Structural Parameter Optimization Pseudo Code

Algorithm: Classification of Hyperspectral Image based on pixel level. The training sample, which has the same size, is cut from the source image. Its corner pixel is regarded as the label.

Input: BS: The batch_size of the samples (the number of samples in each training);
Epoch: A hyperparameter, the maximum number of epochs to train the network;
train_each_epoch_updt: In each training epoch, the number of parameters' updating;
test_each_epoch_pre: In each test epoch, the number of predictions to samples;
X_{train}: The clipped images for training;
X_{test}: The clipped images for testing.
Output: MA: The average accuracy on the test samples;
The trained model's weights
for epoch in range Epoch, do
Training: train the model by training samples
for it in range train_each_epoch_updt do
1) According to the batch extraction method, include the training samples;
2) Acquire average loss by the forward propagation of the network;
3) Update network weights by backward propagation;
end
Testing: Evaluate accuracy for testing samples
for it in range test_each_epoch_pre, do
1) According to the batch extraction method, include the testing samples;
2) Forward propagation;
3) Compute and acquire the accuracy;
end
Compute MA (the average accuracy of all the testing samples).
end

Convolutional layer number and convolution kernel size. This section overviews the influence of the 3D convolutional layer number and the convolution kernel size on the accuracy of the model. Using the univariate principle, the convolutional layer number is first used as a variable to attempt different depths for the network structure. Compare 2, 3 ,4, 5, 6 models of convolutional layers, Under the same number of iterations, the recognition accuracy of the multi-volume layer network model was not significantly improved; however, the computational cost is increased, compared to models of 2 or 3 convolutional layers. The network model of 2 or 3 convolutional layer is more cost-effective. Different convolution kernel sizes were then attempted; the smaller convolution kernel size yielded better results in comparisons of the 3×3×5 convolution kernels on the UAV-HSI data. Because the number of convolution layers and the size of the convolution kernel have little effect on the accuracy of the model, these are not described in great detail here.

Convolution kernel. The convolution kernel is also called a filter. The more convolution kernels there are, the more complete the implicit feature extraction of the sample is. While a model with more kernels can more accurately fit the sample, too many convolution kernels require too many parameters, which slows the training process.

In this section, the conclusions obtained in the previous section are applied to train 3D-CNN network models with different numbers of kernels. Accordingly, the convolution kernel size 3×3×5 is used. C1 and C2 are attempted as 2-layer convolutional network models, and C3, C4, C5, as 3-layer convolutional network models. For C1 and C2 the number of cores is set to 1:2 and 1:3, respectively. For C3:C4:C5 the kernel ratio is set to 1:2:3 and 1:2:4. The model accuracy test was performed separately, and the results are shown in Fig. 4.

As seen in Fig. 4a, when the 3D-CNN has a double convolution kernel, the OA of the 1:2 number of convolution kernel ratio is higher than that of the 1:3 number of convolution kernel convolution ratio. When the number of first-layer convolution kernels is 5 and the second-layer convolution kernel is 10, the OA is the highest, at 92.965%.

As seen in Fig. 4b, when the 3D-CNN has a double convolution kernel, the OA of the 1:2:4 number of convolution kernel ratio is higher than that of the 1:2:3 number of convolution kernel convolution ratio. When the number of convolution kernels in the first layer is 15, the convolution kernel in the second layer is 30, and the number of convolution kernels in the third layer is 60, the OA is the highest at 93.056%. Thus, the overall classification accuracy of the network models with three convolutional layers is superior to that of models with two convolutional layers. For the three convolutions layers, the classification effect is best when the convolution kernel ratio is 15:30:60, returning an OA value of 93.05%.

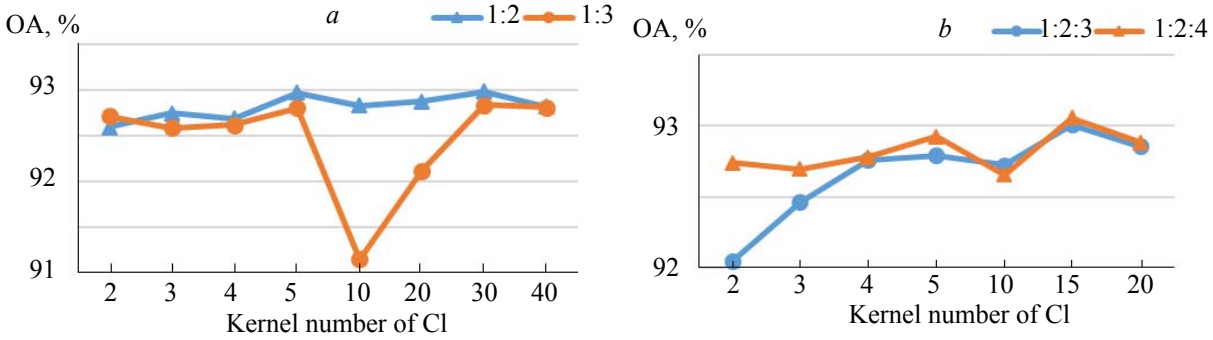

Fig. 4. Overall accuracy of the double (a) and three (b) convolutional layers with different convolution kernels.

Spatial size. Spatial size refers to the size of the image being cropped. In the process of cropping, a larger spatial size often contains more texture information. If the texture of an object is more detailed and contained in a smaller image, then the convolution check will be more focused on the characteristics of small pictures, and may be better than large pictures.

Accordingly, for the experiments, spatial sizes of 7, 9, ..., 15 in an order of 2 are selected, and the test set accuracy and time consumption of the models compared. As the spatial size increases, the training time also increases, while the OA tends to decrease overall (Fig. 5a). When the patch size of 7 has the lowest time cost and highest precision, making the first choice, with an OA value is 94.16%.

Batch size. Batch size is the number of samples used in every training instance of the model. The larger the sample size, the better the ability to represent the overall sample. Thus, in a model with a larger batch size, it is easier to learn the characteristics of the overall sample in every training instance. The parameter update to the global optimization direction provides significant guidance; however, an excessive batch size overloads the memory state. Then, because the number of iterations is too small, the time to achieve the same precision greatly increases, and the parameter correction speed decreases.

Seven different size batch sizes were attempted: 16, 32, 64, 128, 256, 512, and 1024. The results are shown in Fig. 5b. As the batch size increases, the OA tends to decrease. Additionally, the time required to train the model decreases significantly as the batch size increases. This case, the optimal batch size is 128, with an OA value of 96.19%.

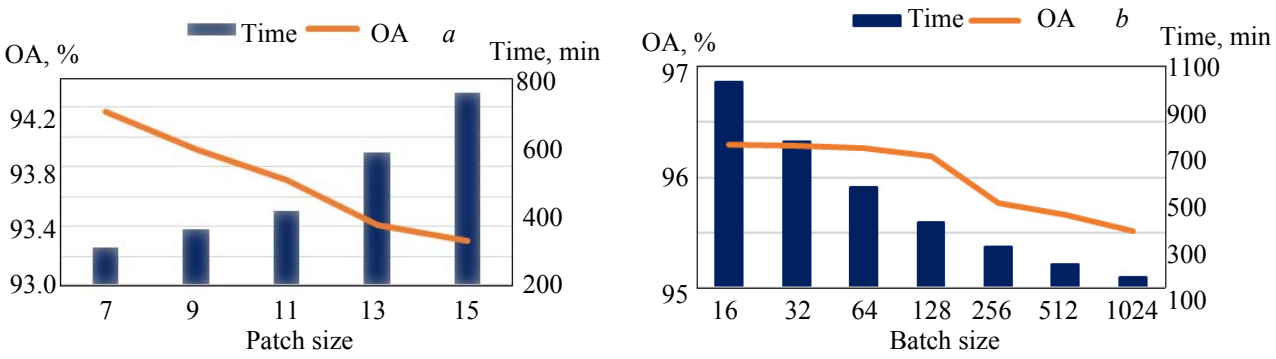

Fig. 5. Accuracy and time consumption (a) of different spatial size and (b) for varying batch sizes.

Other parameters. Through the experiments in Sections 4.2–4.4, the OA of the model was improved from 89.03 to 96.19%. After these experiments, it was found that the activation function, optimization algorithm, learning rate, and training sample have minimal effects on the accuracy of the model.

Four activation function were attempted: Sigmoid, ReLU, ELU, and tanh, of which ReLU showed the best performance. Four optimization algorithms were attempted: SGD, Adam, RMS prop, and Mini-Batch Gradient Descent (MBGD). Among these, Adam showed the best performance. We tried 7 different learning rates, which were 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005, 0.0001, when the learning rate was 0.0005 optimum performance. Eight gradient training samples were attempted for model training, from 10–80% in increments of 10%; the model achieved the best training effect when the training sample was 50%. A 100 epochs model training was conducted. When the epoch is 46, the training precision of the model tends to a stable value, so the epoch takes 50.

Results and discussion. From the above experiments, the three parameters that have the greatest impact on the classification accuracy of the model were determined to be the number of convolution kernels, the spatial size, and the batch size. By optimizing the nine parameters, the OA of the DGC-3D-CNN model was improved to 98.11, 9.14% greater than the OA of the initial model. The specific structure and parameters of the optimized model are Cov1-3×3×5@15, Cov2-3×3×5@30, Cov3-3×3×5@60, ReLU for the activation function, Adam for the optimization algorithm, 0.0005 for the learning rate, and 50% for the training sample.

The original map and the optimized DGC-3D-CNN model classification effect diagram are shown in Figs. 6a,b. To further compare the classification effect, the partial image is enlarged, as shown in Figs. 6c,f. A comparison of the Figs. 6c,d shows the small amount of misclassification at the boundary; it can also be seen from a comparison of the Figs. 6e,f that there is a small amount of misclassification at the boundary of the "other" feature type.

The above experiments result in the following conclusions. The low-altitude UAV hyperspectral remote sensing platform constructed by this experiment has a spatial resolution of 2.1 cm/px, which provides a material basis for accurate identification of small objects. Although the 3D convolution data can be analyzed without data dimension reduction, data reduction with the F-norm used herein can reduce the data dimension while maintaining the spectral spatial information integrity, thus improving the model's training efficiency. The 3D-CNN achieved satisfactory performance in joint spatial spectral information analysis; however, its structure and parameters influence recognition accuracy. After experiments, the number of convolution kernels, batch size, and spatial size greatly influence the accuracy of the network model; with a lighter network structure, a smaller convolution kernel size has a better effect on the identification of small desertified grassland features. After optimization of the nine parameters, the best performance DGC-3D-CNN model was obtained.

Conclusions. This study constructs a low-altitude UAV hyperspectral remote sensing platform, collects data from the Gegentala grassland in Inner Mongolia, and establishes a DGC-3D-CNN desertification grassland vegetation coverage model. The F-norm function is used to reduce the dimensionality of the data. In addition to maintaining spectral characteristic integrity, the training time is effectively shortened. By optimizing the structure and parameters of the DGC-3D-CNN, the best DGC-3D-CNN for a low-altitude drone hyperspectral remote sensing platform for collecting desertification grassland vegetation characteristics is obtained. The optimized model improves the accuracy by 9.8% compared to the initial model, with an overall recognition accuracy of 96.16%. Accordingly, this study provides a method for data collection and analy-

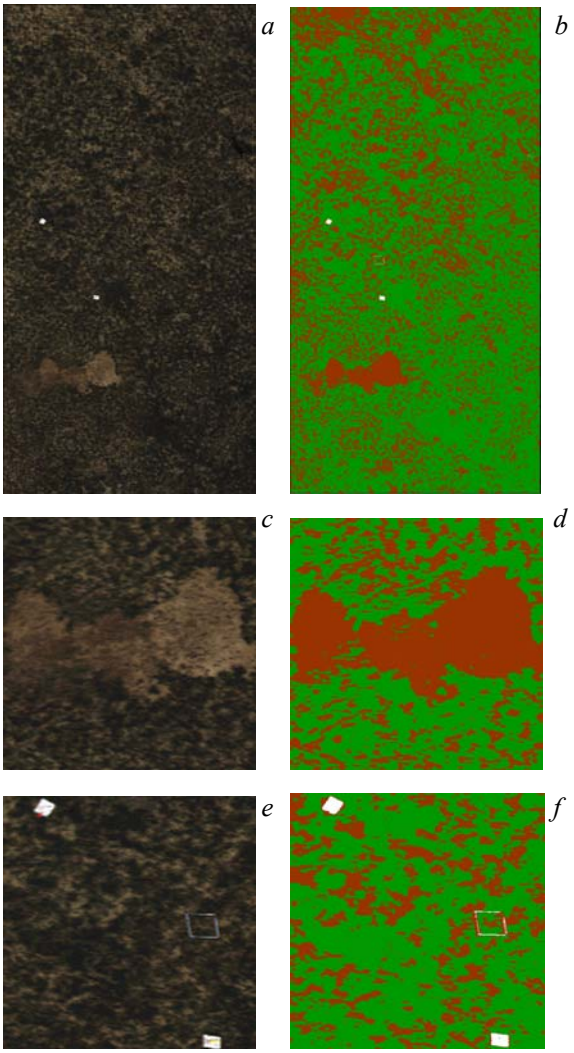

Fig. 6. (a) UAV-HSI, (b) classification result, (c) local image of vegetation and soil, (d) local image of vegetation and soil classification result, (e) local image of vegetation, soil, and other objects, and (f) local image of vegetation, soil, and other objects classification result.

sis of desertification grassland using a UAV hyperspectral remote sensing system. This provides a theoretical basis for hyperspectral remote sensing inversion of vegetation coverage in desertification grassland, while also informing potential further investigations in fine ground object recognition.

Acknowledgments. This study was financially supported by the National Natural Science Foundation of China (No. 31660137). Weiqiang Pi was the primary contributor of this manuscript; he was mainly responsible for data analysis and model optimization and has written the first draft of this manuscript. Hao Liu established the initial model for this study. Xiangbing Zhu contributed to the editing of the manuscript and improved the illustrations.

REFERENCES

1. Q. G. Zhao, G. Q. Huang, Y. Q. Ma, *Acta Ecol. Sin.*, **36**, N 19, 6328–6335 (2016).
2. Q. M. Pan, J. G. Xue, T. Jin, *Chin. Sci. Bull*., **63**, N 17, 1642–1650 (2018).
3. Y. F. Bai, Q. M. Pan, Q. Xing, *Chin. Sci. Bull*., **61**, N 2, 201–212 (2016).
4. Y. Yan, Y. F. Chen, G. C. Zhao, *Geol. Exploration*., **55**, N 2, 630–640 (2019).
5. D. Han, H. Z. Wang, B. Y. Zheng, F. Wang, *Acta Ecol. Sin.,* **38**, N 18, 6655–6663 (2018).
6. Z. H. Gao, B. P. Sun, G. D. Ding, *J. Desert Res*., **84**, N 1, 19–24 (2017).

7. X.-Q. Wei, X.-F. Gu, Q.-Y. Meng, T. Yu, K. Jia, Y.-L. Zhan, Ch.-M. Wang, *J. Appl. Spectrosc.,* N 5, 829–836 (2017).
8. D. Tuia, C. Persello, L. Bruzzone, *IEEE Geosci. Rem. Sens. Magn*., **4**, 41–57 (2016).
9. R. R. Wan, P. Wang, X. L. Wang, *J. Appl. Rem. Sens*., **12**, N 4, 046029 (2018).
10. Q. L. Niu, H. K. Feng, G. J. Yang, *Trans. Chin. Soc. Agric. Eng.*, **34**, N 5, 73–82 (2018).
11. C. Gevaert, J. Suomalainen, J. Tang, *IEEE J. Sel. Top. Appl. Earth Obs.*, **8**, 3140–3146 (2015).
12. G. E. Hinton, N. Srivastava, A. Krizhevsky, *Neural Comput.*, **18**, N 3, 1527–1554 (2006).
13. L. M. Dang, S. Hassan, *Expert Syst. Appl.*, **9**, N 1, 156–168 (2019).
14. H. Chen, Y. Sun, X. L. Li, *Neurocomputing*, **9**, N 356, 83–96 (2019).
15. A. Krizhevsky, I. Sutskever, G. E. Hinton, *Commun. ACM*, **60**, N 6, 84–90 (2017).
16. Q. Zou, L. H. Ni, T. Zhang, *IEEE Geosci. Rem. Sens. Lett*., **12**, N 11, 2321–2325 (2015).
17. X. R. Ma, H. Y. Wang, J. Geng, *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 3282–3285 (2016).
18. Y. Li, H. K. Zhang, X. Z. Xue, *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery*, **8**, N 6 (2018).
19. L. Samaniego, A. Bardossy, K. Schulz, *IEEE Trans. Geosci. Rem. Sens*., **46**, N 7, 2112–2125 (2008).
20. J. Li, J. M. Bioucas-Dias, A. Plaza, *IEEE Trans. Geosci. Rem. Sens*., **48**, N 11, 4085–4098 (2010).
21. Y. Li, H. K. Zhang, Q. Shen, *Rem. Sens*., **9**, N 1, 67 (2017).
22. J. Yue, S. J. Mao, M. Li, *Rem. Sens. Lett*., **7**, N 9, 875–884 (2016).
23. Y. Chen, X. Zhao, X. Jia, *IEEE J. Sel. Top. Appl. Earth Observ. Rem. Sens*., **8**, N 6, 2381–2392 (2015).
24. Y. S. Chen, Z. H. Lin, X. Zhao, *IEEE J. Sel. Top. Appl. Earth Observ. Rem. Sens*., **7**, N 6, 2094–2107 (2014).
25. X. F. Liu, Q. Q. Sun, Y. Meng, *Rem. Sens*., **10**, N 9, 1425 (2018).
26. Y. X. Jin, F, Liu, J. Zhang, *Chin. J. Plant Ecol*., **42**, N 3, 361–371 (2018).
27. Z. B. Xie, P. Wu, G. D. Han, *J. Agric. Mech. Res*., **35**, N 2, 189–191, 196 (2013).
28. S. W. Ji, W. Xu, M. Yang, *IEEE Trans. Pattern Anal. Mach. Intel.,* **35**, 221–231 (2013).
29. W. Zhao, H. Zhang, *Proc. 2012 Int. Conf. Computer Science and Electronics Engineering (ICCSEE 2012)*, 23–25 March 2012, 88–391 (2012).