# HYPERSPECTRAL IMAGE CLASSIFICATION IN DESERT GRASSLAND BASED ON THREE-DIMENSIONAL DEEP LEARNING MODEL

## *基于 3D 深度学习模型的荒漠草原高光谱图像分类*

**Ronghua WANG [1], Yanbin ZHANG [2]*, Jianmin DU [2], Yuge BI[2]**
[1] Inner Mongolia Technical College of Mechanics and Electrics, Hohhot, China
[2] Inner Mongolia Agricultural University, Hohhot, China
*Tel: +86 13214055530; E-mail: zyb359@126.com*
*Corresponding author: Yanbin ZHANG*
*DOI: https://doi.org/10.35633/inmateh-69-46*

***Keywords:*** *hyperspectral remote sensing, unmanned aerial vehicle, deep learning, classification, desert grassland*

**ABSTRACT**

*Identification and classification of vegetation are the basis for grassland degradation monitoring, classification and quantification studies. Here, four deep learning models were used to classify the* unmanned aerial vehicle *(UAV) hyperspectral remote sensing images of desert grassland. VGG16 and ResNet18 achieved better image classification results for vegetation and bare soil, whereas three-dimensional (3D)-VGG16 and 3D-ResNet18, improved by 3D convolutional kernels, achieved better classification for vegetation, bare soil and small sample features in the images. The number of convolutional kernels, its size and batch size parameters of each model were optimised, and 3D-ResNet18-J had the best classification performance, with an overall classification accuracy of 97.74%. It achieved high precision and efficiency in classifying UAV hyperspectral remote sensing images of desert grassland.*

**摘要**

*对植被进行识别与分类是草原退化监测、分级和定量化研究的基础。通过四种深度学习模型对荒漠草原无人机高光谱遥感图像进行分类，VGG16 和 ResNet18 对图像中的植被和裸土取得了较好的分类结果，而经过 3D 卷积核改进的 3D-VGG16 和 3D-ResNet18 模型对图像中植被、裸土和小样本地物均取得了较好的分类潜力。对各模型的卷积核数量、卷积核尺寸和 Batch size 参数优化，发现分类性能最佳的模型为 3D-ResNet18-J，总体分类精度达到 97.74%。实现了对荒漠草原无人机高光谱遥感图像的高精度和高效率分类。*

## INTRODUCTION

With the rapid development of computer hardware, computer vision and deep learning have made considerable progress *(Khekare et al., 2022)*. Computer vision has been used in many applications, including object detection, target localisation, target detection and scene classification *(Luca et al., 2022)*. Its misclassification rate for some datasets is lower than that of human eyes. *(Xu et al., 2021)*. Subsequently, deep learning classification methods were gradually introduced into remote sensing image classification, and the early classification stage was mainly based on the spectral feature classification model of remote sensing data with less dimensionality. As the sensor performance improved and the data dimensionality increased substantially, researchers began to explore the classification model using spatial information features. Recently, inspired by video sequence target detection, classification models with joint spectral–spatial feature information were gradually developed for the analysis of high-dimensional remote sensing data *(Zhang et al., 2020)*. Three-dimensional (3D) convolution can extract spectral–spatial feature information simultaneously, which is a powerful tool for extracting detailed features. Deep learning models based on 3D convolution have been successfully applied to the feature classification of widely used remote sensing datasets such as Pavia University, Indian Pines and Botswana *(Liang et al., 2021)*, while researchers have developed and optimised 3D convolution models for the data characteristics and feature characteristics of widely used remote sensing datasets *(Xiao et al., 2019)*.

The remote sensing monitoring of grassland degradation is mainly based on vegetation community characteristics, cover, biomass and soil properties to identify the degradation degree *(Zhou et al., 2019*; *Esteves de Oliveira et al., 2017)*, while the low-altitude remote sensing platform formed by "Unmanned aerial vehicle (UAV) + hyperspectral remote sensing" has the advantages of abundant data sources, wide coverage, time and effort saving and rapid and repeated deployment *(Liu et al., 2018)*. The obtained hyperspectral images have rich spectral dimensional information and high spatial–spectral resolution, which makes them high-quality data sources for studying the composition of target features at a microscopic scale *(Paoletti et al., 2019)*. Presently, the following bottlenecks exist in grassland vegetation classification. For data acquisition, satellite data, UAV + camera, and UAV + multispectrometer are commonly used, and their spatial resolution is limited. For data classification methods, most of the features are classified by calculus operation, vegetation index operation, threshold difference calculations, etc. The desertified grassland land has high similarity among species groups and sparse and low vegetation. This high similarity among species groups, low vegetation sparseness and fragmented patchy distribution in desertified grassland makes identification and classification extremely difficult and requires high-quality data and classification methods.

The spatial resolution of the UAV hyperspectral remote sensing images collected in this research is up to the centimetre level, which is quantitatively different from the widely used remote sensing datasets, and the features are small, fragmented and sparse. Thus, we fully use the spectral–spatial characteristics of the data to further explore, establish and optimise the deep learning streamlined model applicable to the characteristics of the hyperspectral datasets of desert grassland features to realise the efficient, high precision and intelligent classification and statistics of desert grassland features.

## MATERIALS AND METHODS

### Data acquisition

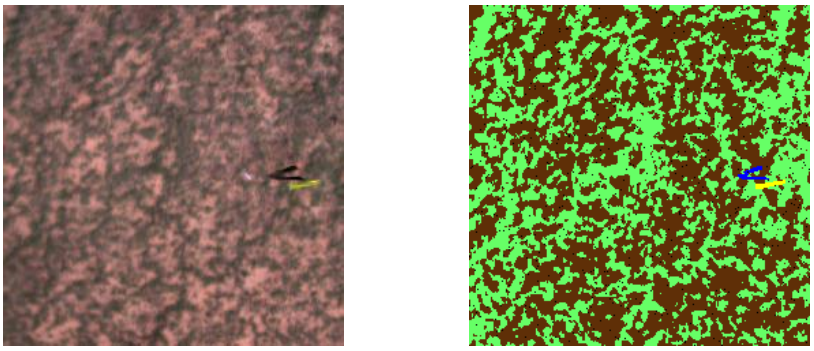
The test area is located in Gegentara Grassland, Siziwang Banner, Ulanqab city (41.78°N, 111.88°E), and a MATRICE 600 PRO six-rotor UAV with a Gaiasky-mini-VN hyperspectral instrument was used to collect hyperspectral remote sensing images of the interior of the test area in hovering mode. The data were collected from July 27 to August 8 during the mushrooming period of forage grass in 2021. To ensure acquisition quality, clear, cloudless days with good light conditions and low wind conditions were selected for the measurement, and the acquisition time was from 10:00 to 14:00. A calibration was required every 10–20 min using a standard whiteboard to eliminate the influence of light intensity changes on the UAV hyperspectral images. The UAV was equipped with a hyperspectral instrument to measure the hyperspectral data of the canopy and the vegetation community in the test area at the height of 30 m perpendicular to the ground directly, and 2–3 hyperspectral remote sensing images were collected at each hovering point. The ground survey included recording GPS information, vegetation species in the occupied area and the bare soil area to obtain accurate ground verification values.

### Data processing

The raw hyperspectral image data captured by Gaiasky -mini-VN were pre-processed in three steps. The first step is image quality evaluation, where the remote sensing images with the best image quality were selected by manually checking to remove overexposed, underexposed, bent and littered images. The second step is radiometric calibration, wherein the original DN values were converted into reflectance using the professional spectral software SpecView. The third step is smoothing and noise reduction, for which the classical Frobenius norm$^2$ method was used to retain the complete spectral space information and achieve data dimensionality reduction *(Wang et al., 2019)*.

### Ground truth value

To improve the calculation efficiency of the four-network model, the collected T1 images of the desert grassland were cropped to a size of 250 rows × 250 columns, with a total of 62,500 image elements to make a hyperspectral dataset, as shown in Figure 1a. The four grassland features in the dataset were labelled as vegetation, bare soil, shadow and others, as shown in Figure 1b. The colour coding and the number of samples are listed in Table 1. Among the labelled samples, 60% were randomly selected as training data, and 40% were test data to identify the features.



*a. False colour composite* *b. Ground truth value*

**Fig. 1 - T1 image false colour composite and the ground truth**

**Table 1**

**Colour coding and sample number of T1 images**

| Class | Colour coding | Number of test samples | Number of training samples | Total |
|---|---|---|---|---|
| **Vegetation** | | 10582 | 15874 | 26456 |
| **Bare soil** | | 14348 | 21524 | 35872 |
| **Shadow** | | 38 | 57 | 95 |
| **Others** | | 30 | 47 | 77 |
| **Total** | | 24998 | 37502 | 62500 |

### VGG16 Network model

Proposed by the Oxford Visual Geometry team, the VGG16 model is a classical neural network model. It is essentially a deep learning approach to improving image recognition and classification accuracy by minimising the loss function in the network by gradient descent and continuously optimising it after several iterations of training *(Zhang et al., 2022).* This model replaces the original 5 × 5 and 11 × 11 convolutional kernels in AlexNet with 3 × 3 convolutional kernels, and the study proves that the 3 × 3 convolutional kernels can reduce the computational complexity, effectively reduce the parameters and improve the training efficiency. The VGG model achieves good results in image classification and image localisation and can effectively extract deep features. In this study, the VGG16 convolutional neural network is built based on the VGG model proposed by Simonyan *(Simonyan et al., 2014)* using the Pytorch module, which has 22 layers, including 13 convolutional layers, five pooling layers, three fully connected layers, and one Softmax layer. The 13 convolutional layers are divided into five convolutional blocks, and a pooling layer is placed after each convolutional block. The minimum number of cores for the convolutional layers is 32, and the maximum number is 1024.

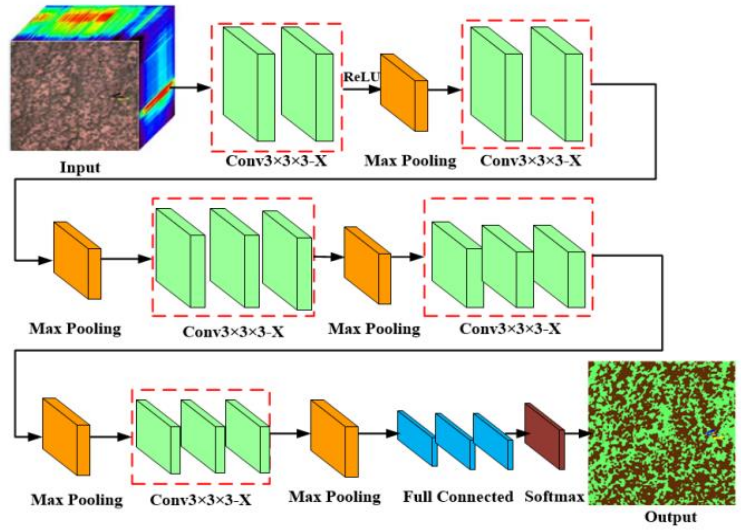
### ResNet18 Network model

Proposed by engineers such as He of Microsoft *(He et al., 2015)*, ResNet18 is a classical neural network model. This model introduces the idea of residual learning into the network structure, superimposes the input and output of the residual block (Residual Block) and sums them through the Skip Connection. This operation improves the efficiency of backpropagation, effectively solves the gradient disappearance and gradient explosion problems that occur in deep networks, accelerates neural network training and realises the use of a deep network structure to extract detailed features *(Zhang et al., 2019)*. The ResNet18 convolutional neural network is built based on the ResNet model proposed by *He et al.* using the Pytorch module, which has 21 layers, including 17 convolutional layers, two pooling layers, one fully connected layer, and one Softmax layer. The 17 convolutional layers are divided into one convolutional layer and eight residual blocks. Two convolutional layer operations are performed in each residual block and superimposed with the parameters before convolution, followed by activation into the next residual block using the activation function ReLU. The minimum number of convolutional layers is 32, and the maximum number is 1024 kernels.

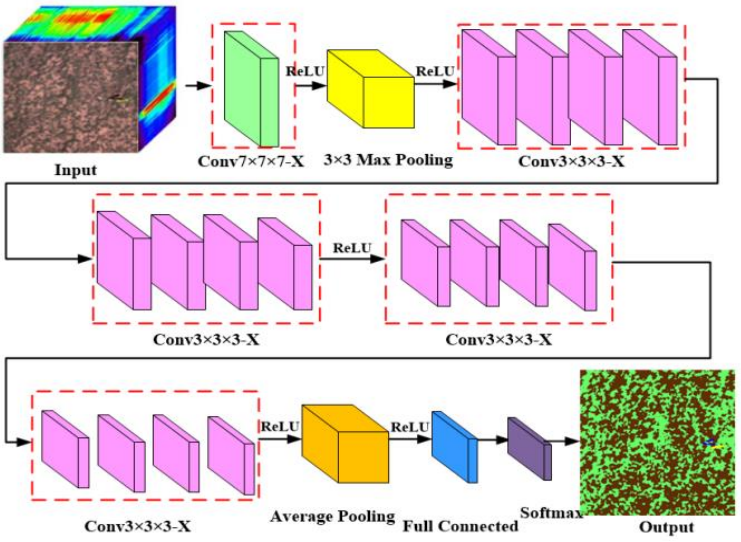
**3D model based on VGG16 and ResNet18**

Based on the Pytorch module, the two-dimensional (2D) convolutional kernels of two typical deep learning models, VGG16 and ResNet18 (hereafter referred to as VGG and ResNet, respectively), were replaced by 3D convolutional kernels to build the 3D-VGG16 model and the 3D-ResNet18 model (hereafter referred to as the 3D-VGG model and the 3D-ResNet model, respectively). Because of the numerous parameters involved in the 3D convolutional operations, the models on a DELL Precision 7920 workstation were trained with the following main parameters: a Xeon Gold CPU containing 16 cores and 32 threads; an NVIDIA Quadro P4000 graphics card with 24 GB of video memory, containing 3840 CUDA cores; and 256 GB of RAM.

3D-VGG comprises 13 convolutional layers (convolutional cores are denoted as Conv1×1×1-X, Conv3×3×3-X, where X is the number of convolutional cores), three fully connected layers (denoted as Fc-x, where x is the fully connected layer level), and five pooling layers (denoted as Max Pooling). 3D-ResNet18 comprises 17 convolutional layers (denoted as Conv3×3×3-X, Conv7×7×7-X, X being the number of convolutional kernels), one fully connected layer (denoted as Fc) and two pooling layers (denoted as Pooling). Two convolutional layer operations are performed in each residual block and superimposed with the parameters before convolution, followed by activation into the next residual block using the activation function ReLU. The 3D-VGG and 3D-ResNet18 deep learning models are structured as shown in Figure 2.




*a. 3D-VGG*




*b. 3D-ResNet*

**Fig. 2 - Structure diagram of the 3D-VGG and 3D-ResNet deep learning models**

**Evaluation indices**

Overall classification accuracy (OA) and producer accuracy (PA) will be used to evaluate the classification accuracy.

(1) The OA refers to the ratio of the number of correctly classified pixels in a random sample to the total number of pixels in the sample, which can reflect the accuracy of hyperspectral remote sensing classification results more intuitively. The formula for OA is shown in equation (1).

$$\mathrm{OA} = \frac{\sum_{i=1}^{k} x_{ii}}{N}, \quad (1)$$

Where:

$N$ represents the number of sample pixels, $k$ represents the number of categories, and $x_{ii}$ represents the composition of class $i$ of the classification result with respect to class $i$ of the reference type data.

(2) The PA refers to the ratio of the number of pixels that correctly classify the entire image into a certain class to the total number of such real samples. The formula for PA is shown in equation (2).

$$\mathrm{PA} = \frac{P_i}{R_i}, \quad (2)$$

Where:

$P_i$ represents the number of samples whose type $i$ is correctly classified in the classification result, and $R_i$ represents the true value of the image element of type $i$.

## RESULTS AND DISCUSSION

The main parameters of the four deep learning models, such as the number of convolutional kernels, kernel size and batch size, were optimised, and the influence of each parameter on the 3D convolutional neural network model was analysed to further improve the accuracy and efficiency of the 3D convolutional neural network model for feature classification.

**Optimisation of the number of convolutional kernels**

The classification performances are compared based on the VGG and 3D-VGG network models with three numbers of convolutional kernels, A, B and C. The corresponding numbers of convolutional kernels are shown in Table 2. The label Conv3×3(×3)-32 represents that 32 convolutional kernels of size 3 × 3 are in the Block_1 layer of VGG-A, 32 convolutional kernels of size 3 × 3 × 3 are in the Block_1 layer of 3D-VGG-A, and so on. The classification performance comparison is based on ResNet and 3D-ResNet network models with three convolutional kernels, A, B and C. The corresponding convolutional kernels are shown in Table 2. In this table, Conv7×7(×7)-32 represents that 32 convolutional kernels of size 7 × 7 are in the Block_1 layer of ResNet-A, 32 convolutional kernels of size 7 × 7 × 7 are in the Block_1 layer of 3D-ResNet-A, and so on.

**Table 2**

**Number of different convolutional kernels for the four models**

| Model | Block_1 | Block_2 | Block_3 | Block_4 | Block_5 |
|---|---|---|---|---|---|
| VGG-A(3D-VGG-A) | Conv3×3(×3)-32 | Conv3×3(×3)-64 | Conv3×3(×3)-128 | Conv3×3(×3)-256 | Conv3×3(×3)-256 |
| VGG-B(3D-VGG-B) | Conv3×3(×3)-64 | Conv3×3(×3)-128 | Conv3×3(×3)-256 | Conv3×3(×3)-512 | Conv3×3(×3)-512 |
| VGG-C(3D-VGG-C) | Conv3×3(×3)-128 | Conv3×3(×3)-256 | Conv3×3(×3)-512 | Conv3×3(×3)-1024 | Conv3×3(×3)-1024 |
| ResNet-A(3D-ResNet- | Conv7×7(×7)-32 | Conv3×3(×3)-32 | Conv3×3(×3)-64 | Conv3×3(×3)-128 | Conv3×3(×3)-256 |
| ResNet-B(3D-ResNet- | Conv7×7(×7)-64 | Conv3×3(×3)-64 | Conv3×3(×3)-128 | Conv3×3(×3)-256 | Conv3×3(×3)-512 |
| ResNet-C(3D-ResNet- | Conv7×7(×7)-128 | Conv3×3(×3)-128 | Conv3×3(×3)-256 | Conv3×3(×3)-512 | Conv3×3(×3)-1024 |

Figure 3 shows that the PA of the VGG model increases with the number of convolutional kernels. The increasing trend in diverse native vegetation and bare soil is more obvious, while the small sample features-

shadow performs best in the VGG-B model. All three models cannot classify the small sample features-other, and the VGG-B model has the best OA.

The 3D-VGG-B model has the best OA at 96.98%, which is 0.52 percentage points better than that of the VGG-B model. 3D-ResNet has better classification performance than ResNet. The PA of the ResNet model increases with the number of convolutional kernels, while the increasing trend in diverse native vegetation and bare soil is smaller. The increasing trend in small sample features-other is more obvious, while the highest accuracy in small sample features-shadow is achieved by the ResNet-B model. The difference in the PA of diverse natives in the 3D-ResNet model is smaller. The best PA was obtained by the 3D-ResNet-B model.

The highest OA, 97.14%, was obtained by the 3D-ResNet-B model and was 2.02% higher than that of the ResNet-B model.

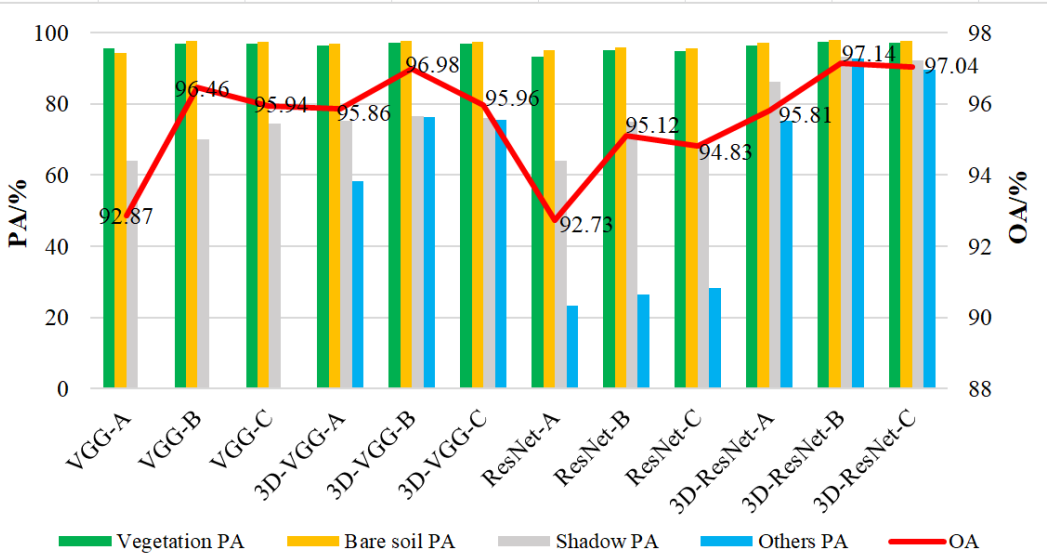



**Fig. 3 - Comparison of the classification performances of models with different numbers of convolutional kernels**

## Convolutional kernel size optimisation

The three convolutional kernel sizes based on the VGG and 3D-VGG network models, identified as D, E and F, are compared in terms of classification performance. The corresponding number of convolutional kernels is shown in Table 3, and the label Conv3×3(×3)-64 represents that 64 convolutional kernels of size 3 × 3 are in the Block_1 layer of VGG-D, 64 convolutional kernels of size 3 × 3 × 3 are in the Block_1 layer of 3D-VGG-D, and so on. The classification performance comparison based on ResNet and 3D-ResNet network models with three convolutional kernel sizes, D, E, and F, and the corresponding number of convolutional kernels are shown in Table 3. Conv7×7(×7)-64 in this table represents that 64 convolutional kernels of size 7 × 7 are in the Block_1 layer of ResNet-D, 64 convolutional kernels of size 7 × 7 × 7 are in the Block_1 layer of 3D-ResNet-D, and so on.

**Table 3**

**Convolutional kernel sizes for the four models**

| Model | Block_1 | Block_2 | Block_3 | Block_4 | Block_5 |
|---|---|---|---|---|---|
| VGG-D(3D-VGG-D) | Conv3×3(×3)-64 | Conv3×3(×3)-128 | Conv3×3(×3)-256 | Conv3×3(×3)-512 | Conv3×3(×3)-512 |
| VGG-E(3D-VGG-E) | Conv5×5(×5)-64 | Conv5×5(×5)-128 | Conv5×5(×5)-256 | Conv5×5(×5)-512 | Conv5×5(×5)-512 |
| VGG-F(3D-VGG-F) | Conv7×7(×7)-64 | Conv7×7(×7)-128 | Conv7×7(×7)-256 | Conv7×7(×7)-512 | Conv7×7(×7)-512 |
| ResNet-D(3D-ResNet-D) | Conv7×7(×7)-64 | Conv3×3(×3)-64 | Conv3×3(×3)-128 | Conv3×3(×3)-256 | Conv3×3(×3)-512 |
| ResNet-E(3D-ResNet-E) | Conv7×7(×7)-64 | Conv5×5(×5)-64 | Conv5×5(×5)-128 | Conv5×5(×5)-256 | Conv5×5(×5)-512 |
| ResNet-F(3D-ResNet-F) | Conv7×7(×7)-64 | Conv7×7(×7)-64 | Conv7×7(×7)-128 | Conv7×7(×7)-256 | Conv7×7(×7)-512 |

The classification performance of the 3D-VGG model with different convolutional kernel sizes is better than that of the VGG model, and all models reach the best classification performance in the F model. The PA of the VGG model is positively correlated with the convolutional kernel size, among which the increasing trend in diverse native vegetation and bare soil and small sample features-shadow is more obvious. In contrast, all three models cannot classify small sample features-other, and the highest OA is that of the VGG-F model. The 3D-VGG model has the highest OA at 97.29%, which is 0.51 percentage points higher than that of the VGG-F model. The 3D-VGG model is positively correlated with the size of the convolutional kernel, and the increasing trend in the small sample features-other is the most obvious.

In terms of classification performance, the ResNet model outperformed the ResNet model, and both models reached their best classification performance in the F model. The PA of the ResNet model was positively correlated with the number of convolutional kernels and showed less increase for diverse features of vegetation and bare soil. In contrast, the PA of small sample features-other increased more, and the ResNet-F model had the highest OA. The PA of the 3D-ResNet model is positively correlated with the number of convolutional kernels, smaller for two diverse natives and larger for two small sample features.

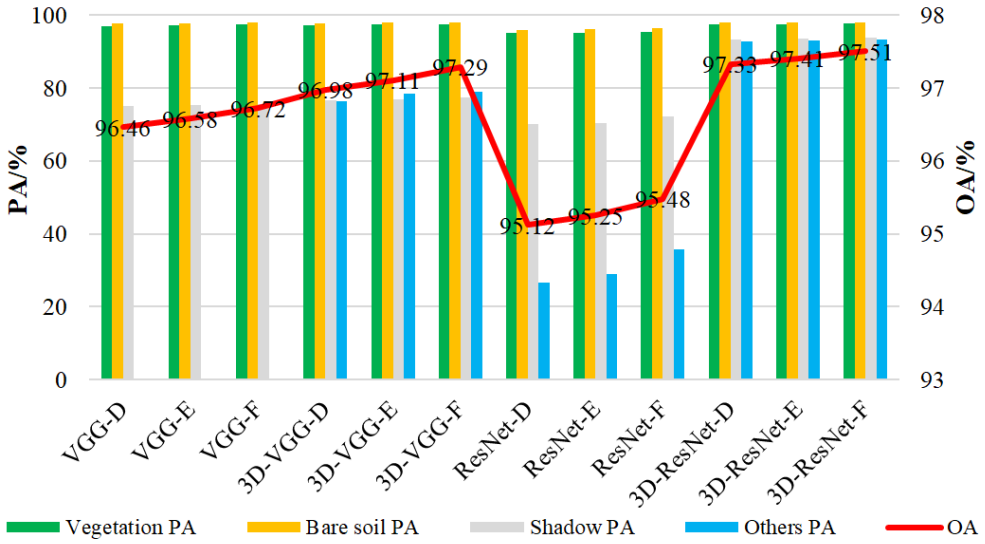



**Fig. 4 - Comparison of classification performances under different convolutional kernel sizes**

## Batch size optimisation

The classification performances of different batch size models on the dataset were investigated for the batch sizes 32, 64, 128 and 256, identified as G, H, I and J, respectively. The corresponding names of the batch size models are shown in Table 4.

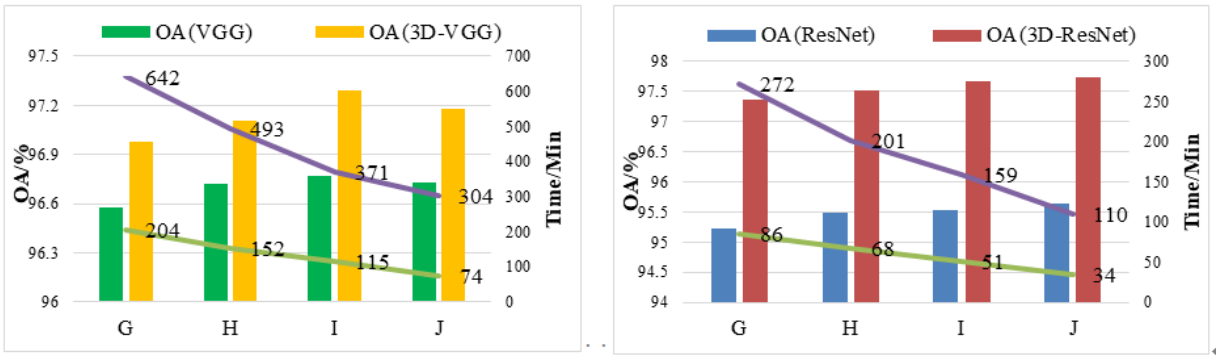
**Table 4**

**Different batch sizes for the four models**

| Batch Size | 32 | 64 | 128 | 256 |
|---|---|---|---|---|
| VGG | VGG-G | VGG-H | VGG-I | VGG-J |
| 3D-VGG | 3D-VGG-G | 3D-VGG-H | 3D-VGG-I | 3D-VGG-J |
| ResNet | ResNet-G | ResNet-H | ResNet-I | ResNet-J |
| 3D-ResNet | 3D-ResNet-G | 3D-ResNet-H | 3D-ResNet-I | 3D-ResNet-J |

Figure. 5a shows that as the batch size gradually increases, the time to reach convergence and the overall classification accuracy show decreasing and increasing trends, respectively, for the VGG model, and the highest OA is obtained at a batch size of 128 (VGG-I). The 3D-VGG model with four batch sizes shows a decreasing trend in convergence time, and an increasing trend in OA as the batch size gradually increases. The highest OA was obtained at a batch size of 128 (3D-VGG-I). A comparison of the performances of the VGG and 3D-VGG models with four batch sizes shows that the VGG model converges sooner and has lower accuracy. The VGG-I model with the best classification accuracy of 96.85% converges in 115 min, while the 3D-VGG-I model with the best classification accuracy of 97.38% converges in 371 min.

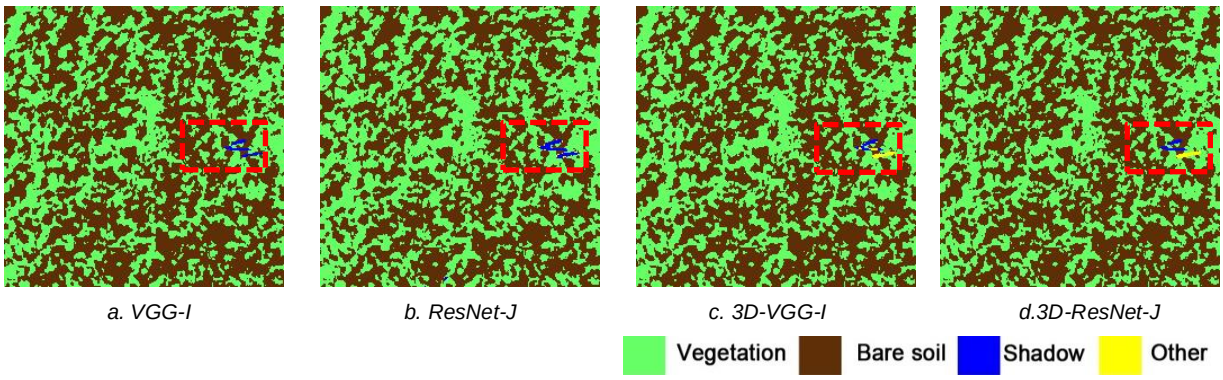
As shown in Figure 5b, as the batch size gradually increases, the convergence times and overall accuracies of the four batch size ResNet models tend to decrease and increase, respectively, and the highest OA is obtained at a batch size of 256 (ResNet-J). The 3D-ResNet models with four batch sizes show a decreasing trend in convergence time and an increasing trend in OA with increasing batch size. The highest OA was obtained at a batch size of 256 (3D-ResNet -J). Comparing the performance of the ResNet and 3D-ResNet models with four batch sizes, it is observed that the ResNet model converges faster and has lower accuracy. The ResNet-J model with the best accuracy of 95.64% converges in 34 min, while the 3D-ResNet-J model with the best accuracy of 97.74% converges in 110 min.




*a. VGG and 3D-VGG* *b. ResNet and 3D-ResNet*

**Fig. 5 - Classification performances of the four models with different batch sizes**

After optimisation, the visualisation results of the best classification effect of the four types of deep learning models are shown in Figure 6.




*a. VGG-I* *b. ResNet-J* *c. 3D-VGG-I* *d.3D-ResNet-J*

**Fig. 6 - Classification visualisation results of the four models**

## CONCLUSIONS

The main conclusions of this study are as follows:

(1) Optimising the model structure and the hyperparameters of the number of convolutional kernels, their size and batch size improves the accuracy of the 2D convolutional model substantially, indicating that the initial 2D model has a considerable improvement space for the feature extraction of small sized features in a dataset. The initial 3D model performs better than the 2D model at dataset feature extraction with less improvement space.

(2) The study found that the 3D convolutional network model has better dataset classification performance than the 2D model. Compared with 2D convolution, which can only extract 2D spectral information in the dataset, 3D convolution can extract combined spectral–spatial information simultaneously, making full use of the data characteristics of hyperspectral datasets and showing good potential for the extraction of vegetation and other features. This potential exists because the 3D convolutional kernel can simultaneously extract spectral information of fine features in the three dimensions of the hyperspectral data than the 2D convolutional kernel that simultaneously extracts spectral information only in two dimensions.

(3) The model performance is gradually improved by optimising the model structure and hyperparameters, and the 3D-ResNet model with a deeper convolutional structure shows better performance. Its producer accuracy and OA are higher than those of the 3D-VGG model. The overall accuracy of the optimised 3D-ResNet-J model reaches up to 97.74%, which is 2–5% higher than the other ResNet models. It achieves high precision, efficiency and intelligent classification of features in the dataset, laying the foundation for fast, efficient and accurate statistics of different vegetation communities in desert grasslands based on UAV hyperspectral remote sensing.

**ACKNOWLEDGEMENT**

The paper has been funded by the National Natural Science Foundation of China (NSFC) (No.31660137) and the Scientific Research Project of the Inner Mongolia Autonomous Region (No. NJZY21518) and special fund support for basic scientific research business expenses of Inner Mongolia Agricultural University (No. BR220152).

**REFERENCES**

[1] He, K., Zhang, X., Ren, S. (2015). Deep residual learning for image recognition. *IEEE*, Vol. 12, pp.1-12.

[2] Khekare, P., Bonthu, S., Hunt, V., Helmicki, A., & Lee, K. (2022). A case study on multilane roundabout capacity evaluation using computer vision and deep learning. *Journal of computing in civil engineering* (3), 36.

[3] Liang, L.H., Li, J., &Zhang, S.Q. (2021). Hyperspectral images classification method based on 3D Octave Convolution and Bi-RNN Attention Network (基于 3D Octave 卷积和 Bi-RNN 注意力网络的高光谱图像分类方法). *Acta Photonica Sinica*, Vol. 50, pp.284-296.

[4] Liu, Z., Wan, W., Huang, J.Y. (2018). Progress on key parameters inversion of crop growth based on unmanned aerial vehicle remote sensing (基于无人机遥感的农作物长势关键参数反演研究进展). *Transactions of the Chinese Society of Agricultural Engineering*, Vol.34, pp.60-71.

[5] Luca, B., Alberto, F., Mauro, J. (2022). Precise Agriculture: Effective Deep Learning Strategies to Detect Pest Insects. *IEEE/CAA Journal of Automatica Sinica*, Vol. 9, pp.246-258.

[6] Paoletti, M.E., Haut, J.M., Plaza, J. (2019). Deep learning classifiers for hyperspectral imaging: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 158, pp.279-317.

[7] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *Computer Vision and Pattern Recognition*, Vol. 9, pp. 22-28.

[8] Esteves de Oliveira, T., David, Santos de Freitas, D., Gianezini, M., Ruviaro, C.F., Zago, D., Mercio, T.Z., Dias, E.A., Lampert, V., Barcellos, J. (2017). Agricultural land use change in the Brazilian Pampa Biome: The reduction of natural grasslands. *Land Use Policy*, Vol. 63, pp.394–400.

[9] Wang, W.G., Song, W., Wang, G.Y. (2019). Image recovery and recognition: A combining method of matrix norm regularization. *IET Image Processing*, Vol.13, pp.1246-1253.

[10] Xu, D., Wang, L., & Li, F. (2021). Review of typical object detection algorithms for deep learning (深度学习的典型目标检测算法研究综述). *Computer Engineering and Applications*, Vol. 57, pp.10-25.

[11] Xiao, Z., Jiang, J., & Ni, C. (2019). Spectral-spatial classification of hyperspectral image based on self-adaptive deep residual 3D convolutional neural network. *Journal of Computer-Aided Design & Computer Graphics*.

[12] Zhang, B., Zhao, L., & Zhang, X. L. (2020). Three-dimensional convolutional neural network model for tree species classification using airborne hyperspectral images. *Remote Sensing of Environment*, Vol.247, pp.1-16.

[13] Zhou, L., Xin, X.P., Li, G. (2019). Application progress on hyperspectral remote sensing in grassland monitoring (高光谱遥感在草原监测中的应用). *Pratacultural Science*, Vol.26, pp.20-27.

[14] Zhang, X.L., Peng, Y. (2019). An audio recognition method based on residual network and random forest (基于残差网络和随机森林的音频识别方法). *Computer Engineering & Science*, Vol.41, pp.727-732.

[15] Zhang, Y.B., Du, J.M., Pi, W.Q., Wang, Y., Gao, X. C. (2022). Deep learning classification of grassland desertification in China via low-altitude UAV hyperspectral remote sensing. *Spectroscopy*, Vol.37, pp.28-35.