GRASSLAND RAT-HOLE RECOGNITION AND CLASSIFICATION BASED ON ATTENTION METHOD AND UNMANNED AERIAL VEHICLE HYPERSPECTRAL REMOTE SENSING
/
基于注意力网络的无人机高光谱草原鼠洞的识别研究

Xiangbing ZHU[1)], Yuge BI[*1)], Jianmin DU[1)], Xinchao GAO[1)], Eerdumutu JIN[1)], Fei HAO[2)] [1]
[1)] Inner Mongolia Agricultural University, College of Mechanical and Electrical Engineering, Inner Mongolia, China
[2)] Department of Mechanical and Electrical Engineering, Hohhot Vocational College, Hohhot, China
Tel:+86-0471-4309215; E-mail: 2538808294@qq.com
DOI: https://doi.org/10.35633/inmateh-70-17

Keywords: *3D convolutional neural networks, Attention mechanism, Rat-hole identification, UAV hyperspectral remote sensing*

ABSTRACT
Rat-hole area and number of rat holes are indicators of the level of degradation and rat damage in grassland environments. However, rat-hole monitoring has consistently relied on manual ground surveys, leading to extremely low efficiency and accuracy. In this paper, a convolutional block attention module (CBAM) model suitable for rat-hole recognition in desert grassland monitoring, called grassland monitoring-CBAM, is proposed that comprehensively incorporates unmanned aerial vehicle hyperspectral remote-sensing technology and deep-learning methods. Validation results show that the overall accuracy and Kappa coefficient of the model were 99.35% and 98.90%, which were 3.96% and 3.35% higher, respectively, than those of the basic model. This study represents a breakthrough in the intelligent interpretation of rat holes and provides technical support for the subsequent rapid interpretation of grassland rat holes and rat damage evaluation. It also provides a solution for the fine classification and quantitative inversion of similar landscape features.

摘要
鼠洞面积和鼠洞数是监测和评价草原退化分级及草原鼠危害等级的双重指标。然而鼠洞监测一直沿用人工地面勘察，效率和精度极低。本研究综合运用无人机高光谱遥感技术和深度学习方法，首次提出了一种适用于荒漠草原监测中鼠洞识别的卷积块注意力模块（CBAM）模型（GM-CBAM）。经过精度验证，模型的总体精度和 Kappa 系数分别为 99.35%、98.90%，相较于基础模型分别提高了 3.96%、3.35%，解决了小样本、高冗余及混合像元导致的识别精度低、泛化能力差等难题，实现了荒漠草原景观下鼠洞的高精度识别。本研究不仅在鼠洞的智能解译方法上有所突破，为后续快速解译草原鼠洞及鼠害等级评价等研究提供了技术支持，也为其他相似景观地物的精细分类和定量反演提供了解决思路。

INTRODUCTION
Grassland rodents not only degrade grassland ecological environments, which restricts animal husbandry production, but also endanger human health. As an important part of the terrestrial ecosystem, grasslands have a practical significance in building an ecological barrier and promoting the animal husbandry economy *(Torok et al., 2021)*. In recent years, owing to the combined effects of global climate change and overgrazing, large-scale grassland degradation has occurred in Inner Mongolia, China *(Lyu et al., 2020)*. Grassland degradation has improved the suitability of the habitat for rodents, resulting in a sharp increase in the rodent population and eventually leading to frequent rodent damage *(Hua et al., 2022)*. The increase in the number of rodents in a grassland environment leads to the consumption of large amounts of vegetation and causes conflict between rodents and livestock. Their tunnelling behaviour causes large areas of secondary bare land, which weakens the self-healing ability of the grassland ecosystem. In addition, rodents carry pathogens that are a serious threat to human and animal health. Therefore, establishing an effective and dynamic grassland rodent monitoring mechanism to protect the grassland ecology and promote animal husbandry production is of great significance.

---
1 *Xiangbing ZHU, Ph.D. Stud. Eng.; Yuge BI, Prof. Ph.D. Eng.; Jianmin DU, Prof. Ph.D. Eng.; Xinchao GAO Ph.D. Stud. Eng.; Eerdumutu JIN Ph.D. Stud. Eng.; Fei HAO, Lecturer M.S. Eng.*

Efficient and accurate monitoring of rodent damage is a prerequisite for devising prevention strategies and measures. Traditional carpet-style manual surveys *(Garba et al., 2014)* are a time-consuming and laborious method of counting rat hole numbers and distribution, and cannot satisfy the demands of large-scale rodent monitoring. Satellite remote sensing has low spatial resolution and cannot acquire detailed spatial distribution information about rat holes. It can only indirectly reflect areas affected by rodent damage through vegetation loss *(Chidodo et al., 2020)*, thus limiting its application. Currently, unmanned aerial vehicle (UAV) remote sensing boasts excellent mobility and efficiency, leading to its wide application in fields such as geological mapping *(Villarreal et al., 2022)*, disaster warning *(Ghali et al., 2022)*, and precision agriculture *(Boursianis et al., 2022)*. The combination of UAVs and hyperspectral imaging enhances the discernibility between rodent damage information on the ground and vegetation shadows *(Qin et al., 2020)*, furnishing the fundamental hardware requirements for identifying rat holes in grassland degradation monitoring. However, further research is necessary to intelligently, accurately, and quickly interpret rat holes from UAV hyperspectral images.

Low recognition accuracy and poor generalization capability are currently bottlenecks in remote sensing classification research. Deep learning is considered the most advanced method in hyperspectral image processing at present. However, these bottlenecks remain unsolved because of the influence of small samples, nonlinear high redundancy, and mixed pixels. Classification methods based on spectral features (1D-CNN) or spatial features (2D-CNN), which ignore the interwoven relationship between space and spectrum, result in insufficient feature extraction capability of the network. Subsequent researchers proposed classification methods based on joint spatial–spectral features (3D-CNN), significantly improving classification accuracy. However, 3D-CNNs are limited in generating large-scale 3D-CNN models because of the number of parameters and computational resources required. Moreover, small-sample training sets are more prone to overfitting issues. Recently, researchers have proposed combining attention mechanisms with CNNs to solve small-sample-size and imbalance issues *(Liu et al., 2022; Wang et al., 2022)*. Attention mechanisms can be understood as an information selection method, focusing on different weight distribution based on the importance of given information and filtering out key information for the current task. Most current research regarding attention remains focused on the classification of urban architecture, forests, and crops, without considering the complexity of desert grassland landscapes. When existing models are applied to the identification of objects in desert grasslands, the recognition accuracy is typically far from expectations because of insufficient generalization ability, particularly in cases with complex topography and fragmented object distribution. Therefore, developing models for fragmented desert grassland landscape objects under the influence of small samples, high redundancy, and mixed pixels and realizing effective feature mining to improve object recognition accuracy is crucial.

To achieve high-precision, intelligent recognition of burrows in complex desert grassland landscapes, in this study hyperspectral data were collected of desert grassland burrow sample plots under natural light using UAV-based hyperspectral remote-sensing technology. By comprehensively employing multiscale 3D convolution, convolutional block attention module (CBAM) attention mechanisms, and dual-branch feature fusion, a GM-CBAM model suitable for burrow recognition in desert grasslands is proposed.

The main contributions of this study are as follows:

1. To address the mixed pixel issue, a grassland monitoring base model (GM-B model) was developed by using algorithms such as multiscale 3D convolution and dual-branch feature fusion. The GM-B model extracts global spatial-spectral joint features with adaptive receptive fields through multiscale 3D convolution, effectively overcoming the interference of mixed pixels. The dual-branch feature fusion enables the feature maps to contain more semantic information while optimizing computational efficiency.

2. To address small-sample-size and data-redundancy issues, attention mechanisms were added to the GM-B model, forming the grassland monitoring CBAM (GM-CBAM) model. The GM-CBAM model filters important features through the attention module, reducing data redundancy and mitigating the overfitting problem caused by small sample sizes.

3. Through performance comparisons between 2D-CNN and 3D-CNN, optimization of parameters such as patch size and learning rate, and performance comparisons and exploration of single-attention modules (channel attention, spatial attention) and mixed-attention modules (CBAM), the influence of channel attention and spatial attention ordering in the CBAM module on the model was investigated. The results enabled the development of an optimal model for grassland burrow recognition.

MATERIALS AND METHODS

Research Area Overview and Data Collection

The study area was in the Gegentala grassland in the central Inner Mongolia Autonomous Region. The area is a typical desert steppe. The constructive species of the grassland is *Stipa breviflora* Griseb, and the dominant species is *Artemisia frigida* Willd. The main rodents are *Lasiopodomys brandtii, Meriones unguiculatus, Rhombomys opimus*, and *Spermophilus dauricus (Zhu et al., 2023)*.

The experimental area covers approximately 3 hm². The spectrometer used for data collection (Fig. 1) has a spectral range of 400–1000 nm, a spectral resolution of 3.5 nm, and 256 bands. At a flight altitude of 20 m, the spatial resolution is 1.73 cm/pixel. Data were collected from July 8 to 10, 2021, with the data collection time window set between 10:00 and 14:00 Beijing time. During this period, solar irradiance was stable, and wind speed was less than 5.4 m/s.

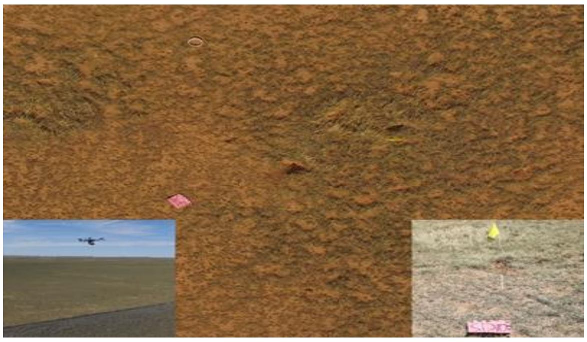

Fig. 1 - UAV hyperspectral data acquisition

Note: Mats and flags in the figure are markers of the rat-hole quadrat.

Data Preprocessing

The raw data were screened for image quality, and distorted and deformed data were removed. After reflectance correction, the standard reflectance images were obtained; the hyperspectral curves for each object are shown in Fig. 2. On the basis of the field investigation, the high-resolution digital images and hyperspectral images were used to create labels for five types of objects: burrows, vegetation, bare soil, sample plot marking foam pads, and sample plot marking flags. The dataset was then divided into training and testing sets at a 4:1 distribution ratio (Table 1).

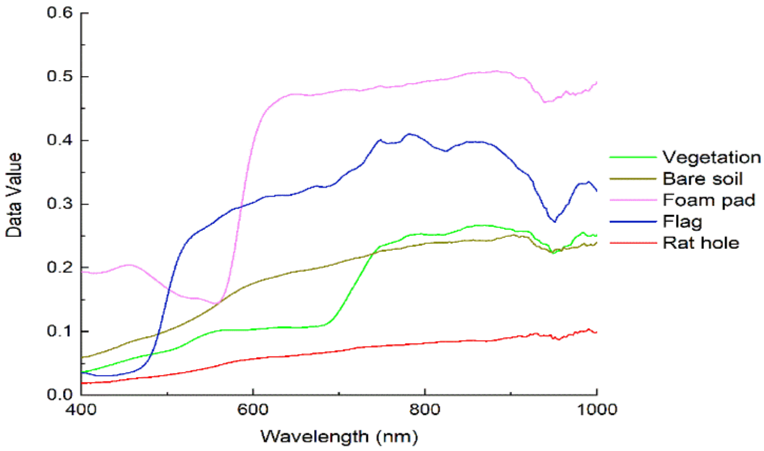

Fig. 2 - Reflectance of different objects

Table 1

Data distribution for training and test sets

NO.	Class name	Training samples	Testing samples
1	Vegetation	9780	2445
2	Bare soil	12220	3056
3	Foam mat	332	83
4	Flag	83	21
5	Rat hole	85	21
	Total	22500	5626

Proposal of GM-CBAM Model

After principal component analysis was applied to the original 256 bands, the first 30 principal components were obtained. To address the mixed-pixel issue, the GM-B model was built using the deep-learning algorithm concept of multiscale 3D convolution and dual-branch feature fusion. To address the small sample size and data redundancy issues a CBAM was added to the GM-B model, forming the GM-CBAM model.

Base Model (GM-B)

The GM-B model first uses 1 × 1 × 1 convolutions for nonlinear enhancement to improve the generalization ability of the model. The model then employs two convolution branches for feature extraction. The first branch uses two layers of 5 × 3 × 3 and 3 × 3 × 3 convolution kernels, whereas the second branch adopts a single layer of 7 × 5 × 5 convolution kernels. Subsequently, the features of the two convolution branches are fused, and finally, a 1 × 1 × 1 convolution operation is applied for nonlinear enhancement again (Fig. 3(a)).

GM-CBAM Model

Built upon the GM-B model, the CBAM is embedded into the convolution branches to form the GM-CBAM model (Fig. 3(b)). The CBAM utilized in this research is essentially a mixed-channel and spatial attention mechanism. Compared to using channel attention mechanisms or spatial attention mechanisms separately, the CBAM module integrates both channel and spatial attention mapping processes, preserving more useful information.

Assuming that the intermediate input feature cube $F \in R^{H \times W \times D \times B}$ is given (where H, W, and D represent the height, width, and spectral band of the feature cube, respectively, and B is the number of channels of the feature cube), the attention weights of different channels are obtained by modeling in the channel dimension, and then the attention weights of different spatial positions are obtained by modeling the feature map F1 (CA output) in the spatial dimension, so as to obtain the dual focus F2 of the target in the channel dimension and the spatial dimension. The mathematical expression of the process is:

$$\begin{cases} F_1 = A_C(F) \otimes F \\ F_2 = A_S(F_1) \otimes F_1 \end{cases} \tag{2}$$

In (2), $A_C(F)$ is the CA feature cube, and $A_S(F_1)$ is the SA feature cube. $\otimes$ is the point multiplication operator, which represents the product of the corresponding elements of the two tensors.

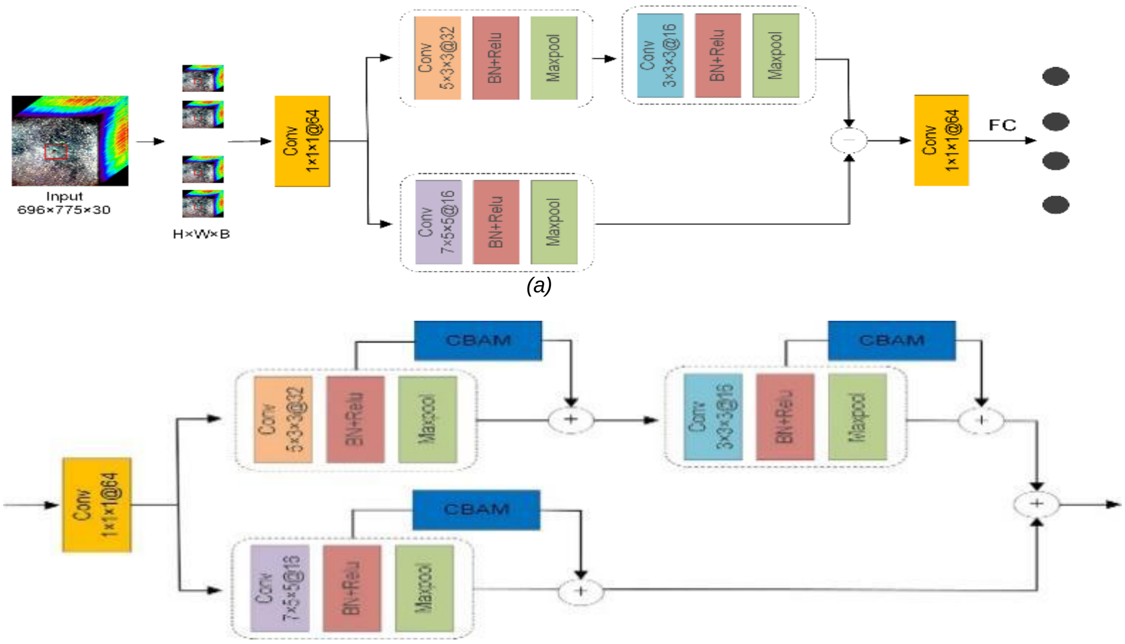

Fig. 3 - Model structure
(a) GM-B model structure; (b) GM-CBAM model

RESULTS AND DISCUSSION

To objectively compare the classification ability of each model, the number of training iterations was set to 100, and the accuracy evaluation index selected the overall accuracy (OA), average accuracy (AA), and Kappa coefficient (K).

Advantages of 3D convolution

To better evaluate the performance of 2D-CNN and 3D-CNN, performance tests under the base network GM-B were conducted. In this test, the patch size was set to 7, the learning rate to 0.001, and the number of fully connected layers to 1024. Figure 4 shows the accuracy results for 2D-CNN and 3D-CNN. Compared with 2D-CNN, 3D-CNN exhibits significant improvements in object recognition rate and OA for all types of objects.

Because hyperspectral images are three-dimensional data, when 2D-CNN is used for classification, the drawback of channel relationship loss exists, preventing 2D-CNN from fully extracting the spectral features required for the classification task. Unlike 2D-CNN, 3D-CNN uses a 3D kernel to extract both spectral and spatial features simultaneously, resulting in features containing more effective and comprehensive high-order semantic information for the classification task. Overall, 3D-CNN demonstrates good classification capabilities in grassland degradation monitoring.

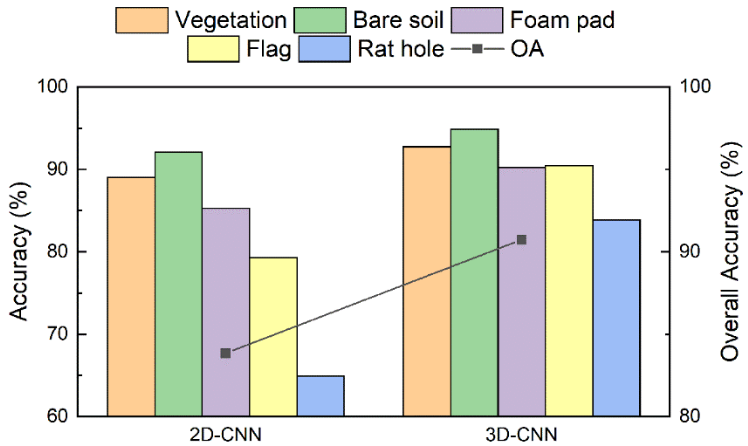

Fig. 4 - Performance comparison between 2D-CNN and 3D-CNN

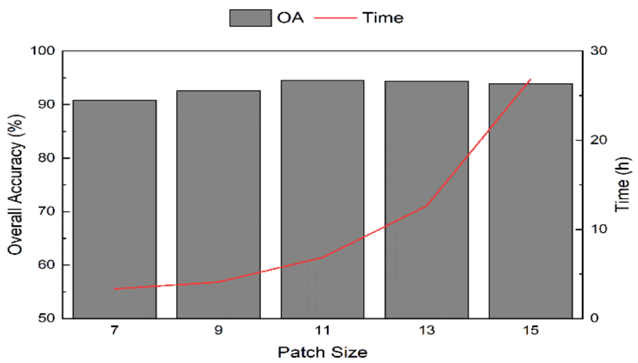

Fig. 5 - Accuracy comparison for different patch sizes

Patch Size Parameter Analysis

The size of the spatial input determines the amount of spatial information in the neighborhood of the central pixel. To evaluate the impact of patch size on model performance, five different patch sizes based on 3D-CNN were set up: 7, 9, 11, 13, and 15. As shown in Fig. 5, the OA value initially increases significantly as the patch size increases, with the best performance resulting when the patch size is 11. When the patch size is greater than 11, the performance improvement is relatively weak, but the training time increases dramatically.

Appropriately increasing the patch size can effectively improve the classification performance of the model, but an overly large patch size may generate substantial noise. In addition, a larger patch size means more training time is required.

Learning Rate Parameter Analysis

Choosing an appropriate learning rate can effectively control the convergence speed and improve the performance of the model. To determine the optimal learning rate, nine different learning rates were set: 0.0001, 0.0003, 0.0005, 0.001, 0.003, 0.005, 0.01, 0.03, and 0.05, on the basis of patch size 11, and the OA values of the model for each learning rate were obtained (Fig. 6). It was found that the model performed best when the learning rate was 0.0005. When the learning rate was greater than 0.005, the performance of the model declined because the larger learning rate made it impossible for the model to reach the global optimum.

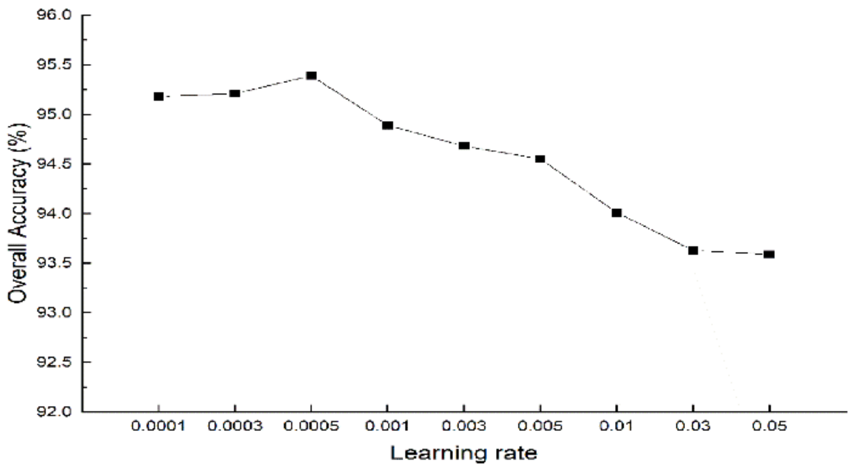

Fig. 6 - Comparison of different learning rates

Comparison of Different Attention Modules and Selection of CBAM Combination

The rat-hole sample is a small sample, and significant imbalance exists among different classes of samples. Moreover, 3D convolution has a large number of parameters. Therefore, attention mechanisms were used instead of data augmentation methods to alleviate the small-sample problem.

To better evaluate the performance of the attention mechanism, the performance of the models were tested including the channel attention (GM-CA), spatial attention (GM-SA), and combined attention (GM-CBAM) modules on the basis of the GM-B model. Three different arrangements are available for the hybrid attention: 1) channel attention prioritized over spatial attention (channel first in CBAM, CBAM_CF); 2) spatial attention prioritized over channel attention (spatial first in CBAM, CBAM_SF); and 3) channel attention and spatial attention in parallel (CBAM_P). As shown in Table 2, all attention modules perform better than the GM-B model, with the hybrid-attention module (GM-CBAM) outperforming the single-attention modules (GM-CA, GM-SA). Among the single-attention modules, GM-CA performs better than GM-SA. Among the three different arrangements in the hybrid attention, CBAM_CF exhibits the best performance, followed by CBAM_P, and CBAM_SF is the worst. At this point, the optimal model GM-CBAM has been obtained; its OA and Kappa coefficient are 99.35% and 98.90%, respectively, an improvement of 3.96% and 3.35% compared to the base model GM-B.

Table 2

Classification results of different models

Class	**Methods**					
#	GM-B	GM-CA	GM-SA	CBAM_CF	CBAM_SF	CBAM_P
Vegetation	98.04	99.94	99.06	100	100	100
Bare soil	98.87	99. 26	99.01	100	100	100
Mat	96.54	99.02	98.93	99.81	99.21	99.30
Flag	90.11	98.07	97.25	99.92	98.94	99.32
Rat hole	90.48	98.47	95.60	98.86	98.58	98.66
OA (%)	95.39	98.93	98.88	99.35	99.13	99.20
AA (%)	95.27	98.49	98.36	99.31	99.03	99.11
K×100	95.55	97.96	96.83	98.90	98.58	98.63

The comparison between the attention models and the base model shows that the attention mechanism, as an information-selection mechanism, has significant advantages in alleviating the small-sample problem and reducing irrelevant information mapping.

Comparing the combined attention module with the single attention modules proves that the hybrid attention mechanism has a significant advantage in mining spectral–spatial feature information in desert grassland hyperspectral images and demonstrates good classification performance.

Comparing the three different arrangements in the hybrid attention shows that channel attention contributes more to model performance than spatial attention. In addition, the result that GM-CA outperforms GM-SA in single attention modules provides further confirmation. From a spatial perspective, channel attention is global, whereas spatial attention is local. Evidently, the weight of local information is determined by the global feature distribution.

Classification Results Display

To visually verify the rationality of the classification results, Fig. 7 shows the visualization results of the original hyperspectral image (Fig. 7(a)), basic network GM-B (Fig. 7(b)), single-attention module GM-SA (Fig. 7(c)), GM-CA (Fig. 7(d)), mixed-attention network CBAM_SF (Fig. 7(e)), CBAM_P (Fig. 7(f)), and CBAM_CF (Fig. 7(g)). Among them, GM-B and GM-SA have noticeable misclassification; the mixed-attention module has no noticeable misclassification, and CBAM_CF exhibits smoother edge control.

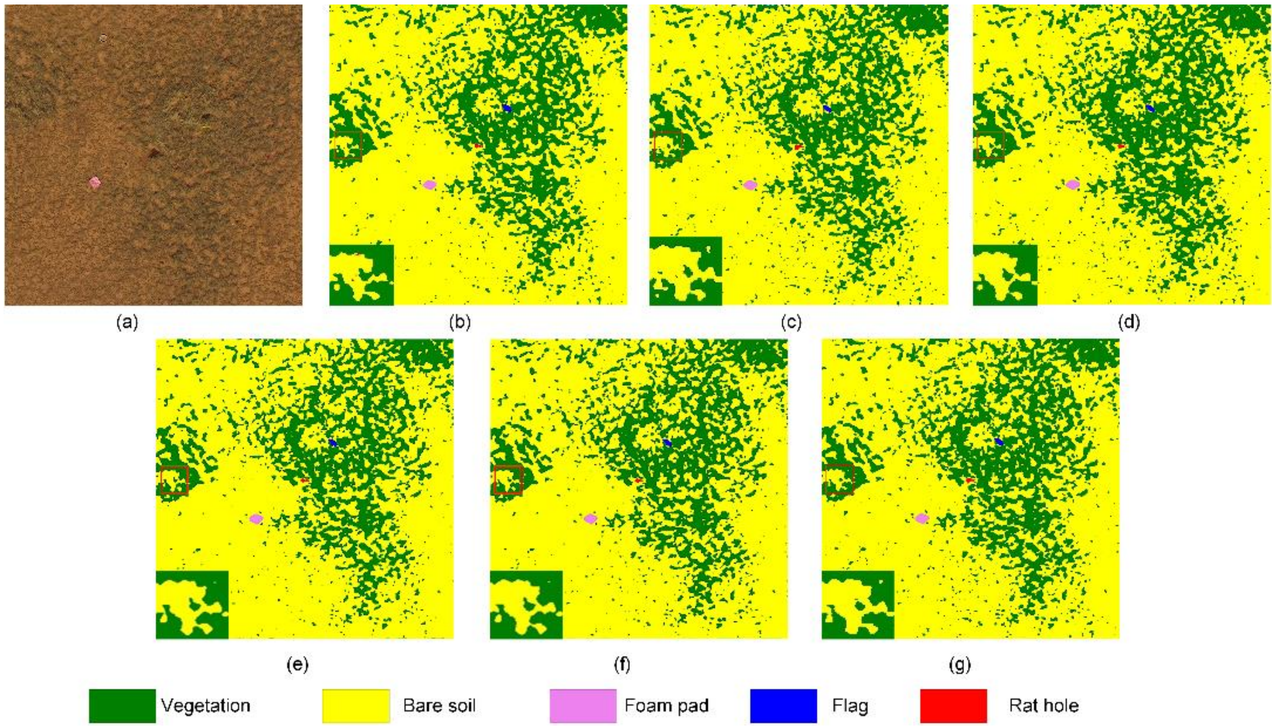

Fig. 7 - Visualization of model classification results

(a) original hyperspectral image, (b) GM-B, (c) GM-SA, (d) GM-CA, (e) CBAM_SF, (f) CBAM_P, and (g) CBAM_CF.
Note: Enlarged images in the boxes in Fig. 7 (b)–(g) are shown in the lower left corner.

Limitations and prospects

The classification of desert grassland landscape features based on UAV hyperspectral images is a complex task, and research in this field is still in the primary stages. Although this study has made acceptable progress, subsequent quantitative inversion and degradation grade evaluation work needs further study. In addition, the GM-CBAM model must be further optimized to enable deployment to mobile devices.

CONCLUSIONS

In this study, a low-altitude UAV hyperspectral remote sensing platform was used to collect desert grassland data in central Inner Mongolia, China. The GM-CBAM model was created with multi-scale convolution, double-branch feature fusion, and a CBAM attention mechanism. After optimizing the network structure and parameters, an optimal model suitable for low-altitude UAV hyperspectral remote-sensing desert grassland classification was obtained. Verification experiments based on the OA, AA, and Kappa coefficient demonstrated the superior classification ability and reliability of the model.

This result provides technical support for the subsequent rapid interpretation of grassland rat holes and rat damage evaluation, in addition to providing methods and models for subsequent quantitative inversion studies of desert grassland.

ACKNOWLEDGEMENT

The research leading to these results received funding from the National Natural Science Foundation of China (grant No. 31660137) and the Research Key Project at Universities of the Inner Mongolia Autonomous Region (grant Nos. NJZZ23037 and NJZY21518) and Inner Mongolia Autonomous Region Natural Science Foundation Joint Fund Project (grant No. 2023LHMS06010).

REFERENCES

[1] Boursianis, A. D., Papadopoulou, M. S., Diamantoulakis, P., Liopa-Tsakalidi, A., Barouchas, P., Salahas, G., & Goudos, S. K. (2022). Internet of Things (IoT) and Agricultural Unmanned Aerial Vehicles (UAVs) in smart farming: A comprehensive review. *Internet of Things*, Vol. 18, Netherlands.

[2] Chidodo, D. J., Kimaro, D. N., Hieronimo, P., Makundi, R. H., Isabirye, M., Leirs, H., & Mulungu, L. S. (2020). Application of normalized difference vegetation index (NDVI) to forecast rodent population abundance in smallholder agro-ecosystems in semi-arid areas in Tanzania. *Mammalia*, Vol. 84, pp. 136–143, Germany.

[3] Garba, M., Kane, M., Gagare, S., Kadaoure, I., Sidikou, R., Rossi, J. P., & Dobigny G. (2014). Local perception of rodent-associated problems in Sahelian urban areas: a survey in Niamey, Niger. *Urban Ecosystems*, Vol. 17, pp. 573–584, Netherlands.

[4] Ghali, R., Akhloufi, M. A., & Mseddi, W. S. (2022). Deep Learning and Transformer Approaches for UAV-Based Wildfire Detection and Segmentation. *Sensors*, Vol. 22, Switzerland.

[5] Hua, L. M., & Chai, S. Q. (2022). Rodents pest control on grassland in China: current state, problems and prospects (中国草原鼠害防治现状、问题及对策). *Journal of Plant Protection*, Vol. 49, pp. 415–423, Beijing/China.

[6] Liu, J. X., Zhang, K. F., Wu, S. Q., Shi, H. T., Zhao, Y. D., Sun, Y. Q., & Fu, E. R. (2022). An Investigation of a Multidimensional CNN Combined with an Attention Mechanism Model to Resolve Small-Sample Problems in Hyperspectral Image Classification. *Remote Sensing*, Vol. 14, Switzerland.

[7] Lyu, X., Li, X. B., Gong, J. R., Wang, H., Dang, D. L., Dou, H. S., & Liu, S. Y. (2020). Comprehensive Grassland Degradation Monitoring by Remote Sensing in Xilinhot, Inner Mongolia, China. *Sustainability*, Vol. 12, Switzerland.

[8] Qin, Y., Sun, Y., Zhang, W., Qin, Y., Chen, J. J., Wang, Z. W., & Zhou, Z. Y. (2020). Species Monitoring Using Unmanned Aerial Vehicle to Reveal the Ecological Role of Plateau Pika in Maintaining Vegetation Diversity on the Northeastern Qinghai-Tibetan Plateau. *Remote Sensing*, Vol. 12, Switzerland.

[9] Torok, P., Brudvig, L. A., Kollmann, J., Price, J. N., & Tothmeresz, B. (2021). The present and future of grassland restoration. *Restoration Ecology*, Vol. 29, United States.

[10] Villarreal, C. A., Garzon, C. G., Mora, J. P., Rojas, J. D., & Rios, C. A. (2022). Workflow for capturing information and characterizing difficult-to-access geological outcrops using unmanned aerial vehicle-based digital photogrammetric data. *Journal of Industrial Information Integration*, Vol. 26, Switzerland.

[11] Wang, L. F., Wang, L. G., Wang, Q. M., & Atkinson, P. M. (2022). SSA-SiamNet: Spectral-Spatial-Wise Attention-Based Siamese Network for Hyperspectral Image Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60, pp. 1–18, United States.

[12] Zhu, X. B., Bi, Y. G., Du, J. M., Gao, X. C., Zhang, T., Pi, W. Q., & Zhang, H. J. (2023). Research on deep learning method recognition and a classification model of grassland grass species based on unmanned aerial vehicle hyperspectral remote sensing. *Grassland Science*, Vol. 69, pp. 3–11, Japan.