UAV image crop classification based on deep learning with spatial and spectral features

Chong Fan*1, Ru Lu1
1Geoscience & Info-Physics, Central South University, Changsha, Hunan, 410083, China
*Corresponding author’s e-mail: fanchong@csu.edu.cn

Abstract. Unmanned Aerial Vehicle image analysis technology has become an effective means to classify crops. However, the UAV image classification mostly adopts shallow learning algorithm, with few computing units. These methods have low recognition accuracy and poor generalization ability. Deep learning has powerful function expression and excellent feature learning abilities. The learned features have more essential character for data and have achieved remarkable results in image classification. Therefore, the study proposes a crop classification method based on Unmanned Aerial Vehicle image with deep learning and spatial and spectral feature fusion. The method uses deep convolutional neural network to classify Unmanned Aerial Vehicle images. Simplified AlexNet network structure, which reduces the number of network layers, was used to accelerate the convergence speed of the model while ensuring the accuracy of crop classification in practical applications. Then, the vegetation index and height features of the Unmanned Aerial Vehicle image were extracted. Feature combination and comparative analyses were carried out to find the most effective feature combination and improve the accuracy of crop classification through spatial and spectral feature fusion. In addition, a Sample Expansion Strategy was used to optimize the classification model and further improve the classification result to achieve a perfect performance in the crop classification of drone images. The experimental results showed that the deep learning method can effectively identify crop types in Unmanned Aerial Vehicle images, and the overall classification accuracy is raised from 86.07% to 92.76% when combining spatial and spectral feature fusion with Sample Expansion Strategy.

1. Introduction
Remote sensing has been used to observe the world in multiple fields, levels, and perspectives. Remote sensing has become an indispensable means of earth observation from the ground to aviation and then to aerospace and from data acquisition, processing, and information extraction to application [1]. Remote sensing is rapid, macroscopic, dynamic, and comprehensive; thus, it is widely used in many fields, among which agricultural monitoring is one of the uses of remote sensing. The remote sensing monitoring system of crop acreage based on the combination of remote sensing and sampling technology has been extensively developed by research institutions in various countries [2]. Sampling technology solves problems in accuracy and timeliness in obtaining planting area information within a regional scope. However, improving the efficiency and reducing the cost of this technology are difficult when conducting a large number of ground surveys. The key to overcome these shortcomings lies in the development of accurate, reliable, and efficient crop classification technology. Satellite remote sensing is the current main data source for crop classification. However, the acquisition of remote sensing images using satellites and aviation sensors with high spatial and temporal resolutions has a long
procurement cycle and high cost, cannot guarantee the time phase, and is easily affected by climate and clouds. Therefore, the data quality obtained cannot meet the requirements. However, unmanned aerial vehicle (UAV) remote sensing with low-altitude imaging technology can make up for the deficiency of traditional satellite remote sensing technology and obtain aerial images with richer spatial information, more obvious ground object information, and higher resolution than satellite remote sensing images. Moreover, UAV remote sensing has low cost, flexible acquisition method, and timely acquisition of crop information on the ground. Classification based on UAV images can be divided into single land cover-type recognition and complex land cover-type recognition.

Gao et al. [3] took the images of soybean in the podding stage and seed filling stage as the research object, then calculated the vegetation index and established univariate and multivariate leaf area index (LAI) inversion models, generated the LAI distribution map of soybean, and reproduced the actual growth situation of soybean. Wang et al. [4] extracted the vegetation information of the study area by constructing the vegetation index of the UAV image and proposed a Visible-band Difference Vegetation Index (VDVI). Compared with other vegetation index, the accuracy of vegetation extraction using vegetation index VDVI is higher, which can effectively extract vegetation information. Ding et al. [5] used four color indexes based on UAV images, namely, Normalized Difference Red–Green Index (NGRDI), Excess Green Index (ExG), Excess Green minus Excess Red Index, and Green Leaf Index, to identify the vegetation coverage area in the study area. The results showed that the recognition accuracies of the four color indexes are more than 90%. Li et al. [6] used binary logistic model to classify the plots; combined the spectral features, texture features, and shape features of UAV images; determined six feature indexes; and realized the extraction of rice planting information. The classification method used is for single land cover type, which uses UAV image analysis technology to extract the spectral and spatial features of the image for vegetation recognition. This kind of research has been more mature and is suitable for single land cover type with obvious difference between foreground and background, but its recognition of complex land cover type is poor. Therefore, machine learning algorithms, such as neural networks and support vector machine (SVM), are applied to the classification of UAV images, and then spectral, texture, and other features are added in the classification process.

Min et al. [7] used back propagation neural network and pixel accumulation, combined the color and texture features of UAV image, and realized typical crop recognition and area measurement with an accuracy of 90.3%. Ahmed et al. [8] used random forest algorithm to classify the land cover types in UAV multispectral camera images with an accuracy rate of up to 95%. Kim and Choi [9] used red–green–blue (RGB) camera and multispectral sensor to obtain UAV images, constructed a digital land surface model and a modified Normalized Difference Vegetation Index (NDVI) map of the study area, and used random forest method to classify land cover types. The results showed that this method has higher classification accuracy than the supervised classification method using only RGB features. Jin et al. [10] used maximum likelihood method and SVM algorithm to classify UAV images to estimate the plant density of winter wheat. The results showed that SVM algorithm has higher classification accuracy than maximum likelihood method. Liu et al. [11] used SVM to classify crops from UAV images and improved the accuracy of crop classification by combining the texture features of UAV images with the height features extracted by the digital terrain model. The results showed that the overall classification accuracy was improved from 72.9% to 94.5% by combining spatial features. The above classification methods are all based on complex ground feature types. Although the classification accuracy has been improved, these classification methods have fewer calculation units, and they will encounter bottlenecks when expressing more complex functions. At the same time, there are some problems, such as long training time, and the generalization ability of the model is not outstanding, which can not meet the needs of crop classification in practical application.

The field of deep learning is developing rapidly. Deep learning networks are composed of multiple nonlinear mapping layers with strong function expression ability and feature learning ability and have been used in large-scale image recognition, object detection, and semantic segmentation with remarkable results [12]-[19]. As a typical deep learning model, convolutional neural networks (CNN),
whose network structure is closer to actual biological neural networks, are highly invariant to translation, scaling, tilting, or other forms of deformation and have achieved great success in image classification [20]. In 2015, Zhao [21] applied CNN to remote sensing image classification for the first time and divided images into three categories: building, bare land, and water; this experiment showed that CNN has higher classification accuracy and shorter classification time compared with traditional classification methods. Cao et al. [22] used high-resolution remote sensing images of Kunming urban area to classify land cover types, including forest land, grassland, house, road, and bare land. They found that the overall classification accuracy of CNN was higher than that of SVM, which further proved the feasibility of CNN in remote sensing image classification. Ji et al. [23] proposed a 3-D CNN method to realize the fine classification of crops by extracting the temporal features of remote sensing images, and they achieved good results. CNN is seldom used in UAV image classification. Li et al. [24] used deep learning and UAV remote sensing technology to integrate a set of hollow village building information survey system. They used CNN to detect the buildings in the hollow village using high-resolution UAV image and achieved good results. Dang et al. [25] introduced deep CNN to classify and evaluate land cover types based on the object-oriented pattern classification system. The results showed that the AlexNet network model can effectively calculate the land type membership of the pattern and realize the automatic quantitative evaluation of land cover type.

Traditional machine learning method is very shallow and relies on the stand or fall of artificial design features; thus, the generalization ability of the model for increasingly complex crop classification problem is not outstanding and could hardly give ideal classification results. In comparison, deep learning network, which is composed of multiple nonlinear mapping layers, has strong function expression ability and obtains good results in complex classification. Therefore, this paper proposes a deep learning method for the fine classification of crops based on UAV and digital surface model (DSM) images. First, an effective and convenient method was selected to calculate the classification features of auxiliary information. Then, CNN was used for model training and classification. Simplified AlexNet network structure was chosen as it guarantees precision while accelerating the convergence rate of the model and shortening training time. A sample was used to expand the strategy based on the classification results after post-processing optimization. Finally, the fine classification of crop in UAV image was realized. The proposed method has a higher overall classification accuracy compared with existing SVM crop classification methods, and the selected classification features are widely used and easy to calculate. Meanwhile, the method has a simple network structure and fast model convergence, which meet the requirements of practical application.

2. Formatting the title, authors and affiliations

This paper aimed to use a popular deep learning method and combine the spectral features of DSM and UAV images to achieve fine crop classification. The technical roadmap shown in Figure 1 consists of the following steps. (1) Selection and combination of classification features. The vegetation index features of the images were extracted and combined with the height features extracted by the DSM. Feature combination and comparative analysis were carried out to find the feature combination that could effectively distinguish crop types. (2) CNN model training and classification. Model training and image classification were carried out using a simplified AlexNet network structure. (3) Classification and post-processing. In case no large-scale misclassification occurred in the classification results, this paper introduced a sample enhancement strategy based on the classification results to obtain more labeled samples and used the enhanced sample data set to train and classify again to optimize the classification model and improve the classification effect. (4) Evaluation of classification results. The overall accuracy and kappa coefficient of the classification results were calculated using confusion matrix as classification evaluation method, and the influences of spectral features, height features, and sample enhancement strategy based on classification results on crop classification accuracy were analyzed. The difference between the proposed method and the current excellent SVM crop classification method was compared.
2.1. Simplified AlexNet CNN model

AlexNet is a classic CNN model. Alex and his tutor Hinton used CNN method in the ImageNet Large Scale Visual Recognition Challenge in 2012 and obtained 10 percentage points higher than the second place using traditional image classification method and won the championship [12]. Deep learning took off afterward, and CNN started to become a household name and grew. The classic AlexNet network model has been used in many applications because of its simple network structure and strong generalization ability, and its application effect has improved with the deepening of the network model [26].
The network has four layers. The first two layers are convolution layers (C1 and C2), and the second two layers are full-connection layers (F3 and F4). We first performed pooling in C1 after obtaining basic convolution data and then performed Rectified Linear Unit (ReLU) and normalization transformation as the input of the next layer. Pooling and normalization transformation were carried out in C2 after ReLU. F3 was connected to C2 and output 384 feature maps. F4 was connected to F3, and the number of output feature maps was 192. Finally, the output of F4 layer is processed with Softmax function to obtain the output. Cross entropy function was used to calculate the output loss of the model during training, and Adam optimizer was used as the optimization method.

2.2. Selection and extraction of classification features
Feature vector selection is one of the decisive factors for classification accuracy. Traditional remote sensing image classification is based on spectral features. However, the same object may have different spectra, and different objects may be present in the same image; thus, an ideal classification result is difficult to attain in complex crop classification problem. Therefore, relying on spectral features is not enough to distinguish the target. The auxiliary information of high-resolution DSM was introduced to finely classify crops in UAV image. Based on DSM, we can distinguish crops from other land cover types by using the elevation of crops, which is different from those of trees, grasslands, and other land cover types. Finally, vegetation index feature and height feature were combined to improve classification accuracy.

The classification method adopted in this paper involved two types of features, namely, vegetation index features and height features, which are described as follows.

2.2.1. Vegetation index features
Vegetation index is a combination of the reflectance of different wavebands, which can effectively reflect vegetation growth and coverage. Therefore, the vegetation index of an UAV image can effectively extract the vegetation information of the image. More than 100 vegetation indices are currently in use in the remote sensing field, but most of them, such as NDVI, are based on visible–near-infrared band. However, UAV image has only three RGB bands excluding the near-infrared band; thus, the vegetation index can only be calculated using the visible band. Few vegetation indexes, including ExG, VDVI, and Red–Green Ratio Index (RGRI), can be calculated using only the visible band. ExG and VDVI have good effect on vegetation information extraction [4-5]. The calculation methods of the two vegetation indexes are shown in Formulas (1) and (2).

\[ EXG = 2 \times \rho_{\text{green}} - \rho_{\text{red}} - \rho_{\text{blue}} \]  
\[ VDVI = \frac{2 \times \rho_{\text{green}} - \rho_{\text{red}} - \rho_{\text{blue}}}{2 \times \rho_{\text{green}} + \rho_{\text{red}} + \rho_{\text{blue}}} \]  

In formulas (1) and (2), \( \rho_{\text{red}}, \rho_{\text{green}}, \text{and} \rho_{\text{blue}} \) are the DN values of the red, green, and blue bands, respectively.

Therefore, ENVI software is used to calculate the ExG and VDVI indices of the UAV image, and the grayscale image of the vegetation index obtained is added to the UAV image as a data band. Then, the combined data are classified.

2.2.2. Height features
Plant height is an important growth index of crops. According to the research findings of other scholars, the combination of height features can be more effective in classifying UAV images [11]. The original color space encounters difficulty in distinguishing the ground features, such as trees, grass, and crops, because of the different height features, introduction of DSM to assist the extraction of height features, and the ability to distinguish ground features.

DSM reflects the shape, height, and other information on the surface of an object. At present, a widely used method to extract height features is to add DSM as a data band to the UAV image through band combination, and then classify the combined data.
2.2.3. SES based on classification results

Post-processing plays a key role in the process of classification. To a certain extent, this step can improve the fragmentation phenomenon in the process of classification and improve the classification accuracy. At the same time, deep learning is a model that simulates the human brain mechanism to analyze, learn, and interpret data. Training of the deep learning model needs a large number of samples, and rich sample datasets can effectively improve the generalization ability of the model. The common methods are image rotation, clipping, and sliding to expand the training data, improve the robustness of the image, and prevent the training process from overfitting. Krizhevsky et al. [12] conducted experiments on an ImageNet dataset and obtained good verification. In the sample extraction process, one of the basic methods is using rotation and inversion to expand the sample data.

In Figure 3, (a) is the original image, and (b, c, d) is the result of rotation of the original image (a) by 90 degrees, 180 degrees, and 270 degrees, respectively. Figure (e, g) is the result of vertical inversion of figure (a, c), and figure (f, h) is the result of horizontal inversion of figure (b, d). In this manner, the training samples can be rapidly expanded to eight times of the number of original samples. Thus, can we optimize the classification model by expanding the sample dataset to perform the post-processing of classification? Based on this idea, the present study proposes a SES based on the classification results for post-processing. In the case of CNN classification without large-area misclassification, more labeled samples are obtained based on the image classification results, and the expanded sample dataset is used for retraining and classification. To a certain extent, this method can reduce the small area spots in the classification process and improve the classification accuracy.

![Figure 3. Simplified AlexNet network structure.](image)

The flow of sample expansion based on classification results is shown in Figure 4. Based on the image classification results, a sliding window is used starting from the upper left corner of the classification result graph. The window size is consistent with the step and sample sizes to ensure that no overlap exists between samples and prevent the redundancy of the sample dataset. Then, the percentage of each category of pixels in the total number of pixels in the window is counted. When the value is greater than the given threshold, the current window can obtain candidate samples, record the current window position and pixel value, and crop the image data to obtain labeled samples. Finally, a new classification model is obtained by retraining the labeled sample set obtained by the SES.
The following inequalities are the design rules for selecting candidate samples:

\[ A < T \]

In formula (3), \( A \) is the percentage of each category of pixels in the window to the total number of pixels, and \( T \) is a given threshold (for example, 0.9). Inequality means that if \( A \) is less than \( T \), then the current window cannot obtain high-quality candidate samples. On the contrary, the current window position is recorded and the image data are clipped to obtain labeled samples.

As the SES proposed in this paper is an automatic process, if the classification effect of CNN is not good and a large area of misclassification phenomenon exists, then the introduction of the SES proposed in this paper will generate a large number of error samples, which will reduce the accuracy of model classification. Therefore, before using the SES based on the classification results, we need to conduct a manual evaluation to determine whether the strategy can be used for post-processing.

3. EXPERIMENT AND ANALYSIS

3.1. Data source
The data used in this study are from the UAV remote sensing test conducted in the experimental area on August 3 and 4, 2017. These data include the UAV image, DSM, and sample point data of Harbin democracy base collected in August 2017. The data format of UAV image and DSM is TIFF, with spatial resolution of 0.1 M and size of 8192 pixels \( \times \) 8192 pixels. The data of 41,000 sample points is in TXT format. The study area consists of 8 types of ground features in the study area, including road, building, harvested land, soybean, corn, potato, mungbean, other, as shown in Figure 5.
According to the research results of reference [11], for the current UAV data, when the 27 pixel × 27 pixel window is used to calculate the variance and inter-class difference coefficient of DSM, the overall classification accuracy of crops is the highest. Therefore, 27 × 27 pixels is considered to be the appropriate sample size for this UAV image. According to the previously collected data of sample points, ENVI IDL software is used for sample clipping. The sample size of 27 × 27 pixels is clipped with the sample point as the center. Then, the sample dataset is made, including the training and verification sets. The number of samples in each category is reported in Table 1.

| Number | Category name | Training dataset | Testing dataset |
|--------|---------------|------------------|----------------|
| 1      | Road          | 2000             | 200            |
| 2      | Building      | 1100             | 200            |
| 3      | Harvested land| 4200             | 300            |
| 4      | Soybean       | 10800            | 1200           |
| 5      | Corn          | 3000             | 400            |
| 6      | Potato        | 5000             | 600            |
| 7      | Mungbean      | 10000            | 1000           |
| 8      | Other         | 900              | 100            |
|        | Total         | 37000            | 4000           |

3.2. CNN model training

The experiment uses TensorFlow, a deep learning framework developed by Google, and CUDA to perform GPU acceleration. Using the simplified AlexNet CNN model, we set the other parameters as follows: each batch 37 samples are taken from the training data set for training. For every 1000 batches of training, we will test the whole test data set, we test 40 samples each batch. The learning rate is fixed at 0.001, momentum is 0.9, weight attenuation is 0.04, learning rate strategy is “fixed,” and the maximum number of iterations is 10,000.

The simplified AlexNet network structure has a good learning effect on the dataset collected in this experiment. When only 4,000 times of training is performed (37 samples each time), the correct rate can reach approximately 91%. The improved network structure has effectively learned the dataset, with fast convergence speed and high accuracy, as shown in Figure 6.
3.3. Feature combination

In evaluating the effective feature combinations, this study conducts experiments to verify them. With the help of ENVI remote-sensing image processing software, the calculated vegetation index and DSM were added to the UAV image in the method of band combination to obtain different features of combination data. Seven feature combinations are selected: (1) using the original RGB features for classification; (2) using RGB features and vegetation index ExG for classification; (3) using RGB features and vegetation index VDVI for classification; (4) using RGB features and DSM for classification; (5) using RGB features and vegetation index ExG; (6) using RGB features and vegetation index VDVI and DSM for classification; and (7) using RGB features and vegetation index VDVI, ExG, and DSM for classification. Then, sample datasets were made for model training and the test accuracy of each feature combination was counted. The results are shown in Table 2.

Table 2. Comparison of classification results of feature combination.

| Number | Category name | Training dataset | Testing dataset |
|--------|---------------|------------------|----------------|
| 1      | Road          | 2000             | 200            |
| 2      | Building      | 1100             | 200            |
| 3      | Harvested land| 4200             | 300            |
| 4      | Soybean       | 10800            | 1200           |
| 5      | Corn          | 3000             | 400            |
| 6      | Potato        | 5000             | 600            |
| 7      | Mungbean      | 10000            | 1000           |
| 8      | Other         | 900              | 100            |
| Total  |               | 37000            | 4000           |

According to comparison and analysis, combination 6 (RGB feature, vegetation index VDVI, and DSM) exhibits the highest test accuracy and can effectively identify various crop types. The test accuracy of the original RGB feature is 89.8%, and the test accuracy is improved after the addition of the ExG or VDVI index due to the spectral difference of various ground features. At the same time, classification accuracy was also improved with the addition of DSM by increasing the height of the information to separate crops from other features. The combination of vegetation index and height features enables the measurement accuracy to reach approximately 95%. For UAV images in the study area, VDVI is more suitable for crop classification than ExG. Table 1 shows that the test accuracy of combination 3 is higher than that of combination 2, and the test accuracy of combination 6 is also higher than that of combinations 5 and 7. The threshold range of VDVI is smaller, near 0, while the threshold
range of ExG is larger. Therefore, combination 7 is greatly affected by the ExG vegetation index, resulting in a decrease in accuracy.

Experimental results show that compared with the original RGB features, the addition of spectral features and height features greatly improves the classification accuracy. Therefore, this study selects the RGB feature, VDVI index, and DSM as the effective feature combination, and comprehensively applies the spectral and height features of the UAV image to improve the classification accuracy.

3.4. Result analysis

To verify the effectiveness of the proposed method, we conducted three groups of experiments to classify the UAV images only containing visible light band and the most effective feature combination (original RGB feature, and vegetation indices VDVI and DSM) obtained previously. Then, the SES based on the classification results was introduced to optimize the post-processing without large-scale misclassification. The results are shown in Figure 7.

![Classification results](image)

When only RGB features are used for classification, the fragmentation of the soybean region and Mungbean region is serious. At the same time, there are also large areas of misclassification such as harvested land divided into Potato and Mungbean divided into harvested land. Combined with vegetation index VDVI and DSM, the classification effect is greatly improved, and fragmentation and misclassification are significantly reduced. Therefore, based on the classification results obtained by RGB, VDVI, and DSM, the SES based on the classification results proposed above is introduced for post-processing optimization. For the UAV image data in this study, the sample size is 27 × 27 pixels, so the sliding window size is 27 × 27 and the step size is 27. To obtain high-quality samples, this study sets the threshold value t to 1, that is, only when all the pixels in the window are classified with the same value can the current window position be determined and the labeled samples be obtained by clipping. After the SES is used, the number of labeled samples is increased from 37700 to 42315. The new
samples are placed in CNN for retraining and classification. The classification results are presented in Figure 7 (d). The classification effect is further improved, the fragmentation of rice and tree areas is reduced, and the shadow misclassification is also improved.

To obtain the objective evaluation results, the accuracy of each classification result, overall classification accuracy, and kappa coefficient are calculated. The results are shown in Table III.

Table 3. Comparison of classification results of feature combination.

| Number | Category name | Training dataset | Testing dataset |
|--------|---------------|------------------|----------------|
| 1      | Road          | 2000             | 200            |
| 2      | Building      | 1100             | 200            |
| 3      | Harvested land| 4200             | 300            |
| 4      | Soybean       | 10800            | 1200           |
| 5      | Corn          | 3000             | 400            |
| 6      | Potato        | 5000             | 600            |
| 7      | Mungbean      | 10000            | 1000           |
| 8      | Other         | 900              | 100            |
| **Total** |               | **37000**       | **4000**       |

Compared with the original RGB feature classification, the fusion of spatial and spectral features greatly improves the classification accuracy. At the same time, the SES can further optimize the classification model and improve the classification results when the classification model performs well. Therefore, the UAV image classification method based on deep learning and spatial spectral features is effective.

3.5. Comparison and analysis

According to the experimental data in the study area, some scholars have used the SVM method to extract the height features of DSM and texture features of the UAV image, and achieved fine classification of crops as well as good classification results [11]. In this study, we use the CNN method combined with the vegetation index features and DSM height features. Then, we introduce the SES based on the classification results, and finally achieve the fine crop classification of UAV images. For comparison, the crop classification method based on reference [11] is compared with the method in this paper. The results are shown in Figure 8.

(a) Classification results of [11]  (b) Classification results of this study

Figure 8. Comparison of classification results.
4. Discussion and conclusion
In this study, we use the popular deep learning method and combine the spatial and spectral features to achieve the fine crop classification of UAV images. Considering the timeliness of crop classification in practical application, we find that the improved AlexNet network structure reduces the number of layers of the model and accelerates the convergence speed of the model on the premise of ensuring a certain accuracy. Then, the spatial spectral features of the image are fused to improve the classification accuracy. Among them, vegetation index VDVI is selected as the spectral feature because the threshold range of VDVI is small (near 0) and has a good extraction effect for rice, buildings, trees, grassland, and others. The spatial feature is introduced into the crop height feature through the fusion of DSM, which facilitates distinguishing the crops that are difficult to distinguish by texture and color. Finally, in the case of good performance of the classification model, this study proposes a new SES to expand the labeled sample set, which further improves the classification results and accuracy. Based on the analysis of the experimental results, combined with deep learning, spatial spectral features, and SES, the fine crop classification of UAV images can be effectively realized.

In the future, we will consider the application of deep learning algorithms with better classification effect, such as deep learning image segmentation technology, to segment UAV images, obtain the corresponding crop categories, and improve the efficiency of large-scale operation. To ensure the efficiency of crop classification in practical application, this study only combines the vegetation index features that are easy to calculate in the UAV images and the height features extracted from DSM data for classification. Many image features still have to be studied so that the distribution of various crops can be accurately identified.

Acknowledgments
This work was supported by the National Key Research and Development Program of China under Grant 2016YFC0803108.

References
[1] Mei, A., Peng, Qin, W., Liu, H. (2001) Introduction to Remote Sensing. Higher Education Press, Beijing.
[2] Wu, B., Meng, J., Li, Q., Zhang, F., Du, X., Yan, N. (2010) Latest development of “CropWatch” - an global crop monitoring system with remote sensing. Advances in Earth Science, 25: 1013-1022.
[3] Gao, L., Yang, G., Wang, B., Yu, H., Xu, B., Feng, H. (2015) Soybean leaf area index retrieval with UAV (Unmanned Aerial Vehicle) remote sensing imagery. Chinese Journal of Eco-Agriculture, 23: 868-876.
[4] Wang, X., Wang, M., Wang, S., Wu, Y. (2015) Extraction of vegetation information from visible unmanned aerial vehicle images. Transactions of the Chinese Society of Agricultural Engineering, 31: 152-159.
[5] Ding, L., Li, Q., Du, X., Cheng, Y., Yuan, C. (2016) Vegetation extraction method based on color indices from UAV images. Remote Sensing for Land & Resources, 28: 78-76.
[6] Li, M., Huang, Y., Li, X., Peng, D., Xie, J. (2018) Extraction of rice planting information based on remote sensing image from UAV. Transactions of the Chinese Society of Agricultural engineering, 34: 108-114.
[7] Min, W., Jiang, C., Li, X., Jiang, G., Rao, Y. (2017) Area measurement for field crops based on UAV platform and image analysis. Journal of Hunan Agricultural University (Natural Sciences), 43: 212-216.
[8] Ahmed, O.S., Shemrock, A., Chabot, D., Dillon, C., Williams, G., Wasson, R., Franklin, S. E. (2017) Hierarchical land cover and vegetation classification using multispectral data acquired from an unmanned aerial vehicle. Int J. Remote Sens., 38: 2037-2052.
[9] Kim, G.H., Choi, J.W. (2017) Land Cover Classification with High Spatial Resolution Using Orthoimage and DSM Based on Fixed-Wing UAV. Journal of the Korean Society of
Surveying, Geodesy, Photogrammetry and Cartography, 35: 1-10.

[10] Jin, X., Liu, S., Baret, F., Hemerle, M., Comar, A. (2017) Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. Remote Sens. Environ., 198: 105-114.

[11] Liu, B., Shi, Y., Duan, Y., Wu, W. (2018) UAV-based Crops Classification with joint features from Orthoimage and DSM data. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 42: 7-10.

[12] Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012) Imagenet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems. New York. pp. 1097-1105.

[13] Simonyan, K., Zisserman, A. (2014) Very deep convolutional networks for large-scale image recognition. https://arxiv.org/abs/1409.1556.

[14] He, K., Zhang, X., Ren, S., Sum, J. (2016) Deep Residual Learning for Image Recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. Nevada. pp. 770-778.

[15] Ranjan, R., Patel, V.M., Chellappa, R. (2019) Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE Trans. Pattern Anal. Mach Intell., 41: 121-135.

[16] Girshick, R., Donahue, J., Darrell, T., Malik, J. (2016) Region-based convolutional networks for accurate object detection and segmentation. IEEE Trans. Pattern Anal. Mach Intell., 38: 142-158.

[17] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016) You only look once: Unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition. Nevada. pp. 779-788.

[18] Long, J., Shelhamer, E., Darrell, T. (2015) Fully convolutional networks for semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. Massachusetts, pp. 3431-3440.

[19] Noh, H., Hong, S., Han, B. (2015) Learning deconvolution network for semantic segmentation. In: IEEE International Conference on Computer Vision, Santiago. pp. 1520-1528.

[20] Dahl, G.E., Yu, D., Deng, L., Acero, A. (2012) Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. IEEE Trans. Audio. Speech. Lang Process., 20: 30-42.

[21] Zhao, S. (2015) Remote sensing image classification method based on convolutional neural networks. M.S. thesis, Dept. Land Science and Technology., China Geosciences Univ.

[22] Cao, L., Li, H., Han, Y., Yu, F., Gu, H. Y. (2016) Application of convolutional neural networks in classification of high resolution remote sensing imagery. Science of Surveying and mapping, 41: 170-175.

[23] Ji, S., Zhang, C., Xu, A., Shi, Y., Duan, Y. (2018) 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. Remote Sens, 10: 75.

[24] Li, Z. (2018) Research on key methods for building information acquisition of hollow village based on high-resolution Image. Ph.D. dissertation, Dept. Faculty of geosciences and environmental Engineering., Southwest Jiaotong Univ.

[25] Dang, Y., Zhang, J., Deng, K., Zhao, Y., Yu, F. (2017) Study on the evaluation of land cover classification using remote sensing images based on AlexNet. Journal of Geo-information Science, 19: 1530-1537.

[26] Wang, D., Fan, C., Mo, D. (2018) Design of remote sensing image crop classification system based on deep learning service. China Agricultural Informatics, 30: 82-92.