This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3090085, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing



# A Lightweight Model of VGG-16 for Remote Sensing Image Classification

Mu Ye,Ni Ruiwen, Zhang Chang,Gong He,Hu Tianli,Li Shijun*,Sun Yu,Zhang Tong ,Guo Ying

*Abstract*—**In planetary science, it is an important basic work to recognize and classify the features of topography and geomorphology from the massive data of planetary remote sensing. Therefore, this paper proposes a lightweight model based on VGG-16, which can selectively extract some features of remote sensing images, remove redundant information, and recognize and classify remote sensing images. This model not only ensures the accuracy, but also reduces the parameters of the model.According to our experimental results, our model has a great improvement in remote sensing image classification, from the original accuracy of 85% to 98% now. At the same time, the model has a great improvement in convergence speed and classification performance.By inputting the remote sensing image data of ultra-low pixels (64 * 64) into our model, we prove that our model still has a high accuracy rate of 95% for the remote sensing image with ultra-low pixels and less feature points.Therefore, the model has a good application prospect in remote sensing image fine classification, very low pixel, less image classification.**

*Key Words*—**VGG-16 Less-feature-points nonlinear-correction-layer zero-padding**

## I. Introduction

REMOTE sensing scence classification has received considerable attention recently,as can be used inmany practical applications,such as natural hazards dection,geographic image retriecal,urban planing,and so on.Given a query remote sensing image, scene classification aims to assign a unique label to the image, based on its contents[1-3].In the early works, handcrafted features are the most widely used in this task and have been intensively investigated, such as color histograms, scale-invariant feature transform (SIFT), and histogram of oriented gradients (HOG). These methods rely heavily on professional skills and domain expertise to design various features so that their adaptability and expression ability are not strong enough.In the meanwhile,remote sensing scene classification is a challenging problem since the scene images often exhibit complex spatial structures with high intraclass and low interclass variabilities. To address this problem, many scene classification methods have been proposed over the past years[4-8] .

Convolutional neural networks (CNN) has a strong ability of generalization and natural image classification. The main reason that CNN model can accurately classify is that it has a large number of label training data sets [9-11].Recently, the application of CNN model has been extended to remote sensing scene classification. However, due to the limited labeled Remote sensing image data, CNN model needs to train these data sets from scratch. In order to solve this problem, transfer learning method will be adopted.Someone used existing remote sensing images for fine-tuning.At the same time, some CNN models (such as AlexNet, Google Net or VGGNet) have trained some large-scale data in advance [12].Generally speaking, it has been proved to be effective to fine tune the convolution layer of a pre trained CNN model to adapt the architecture to new classification tasks.Compared with traditional classification methods, these models have better classification performance. The success of CNN based scene classification method is mainly due to the strong generalization ability of the pre trained CNN model, which can extract more representative features than traditional feature extraction methods [13].

Deep learning proves very promising for land use and land cover classification (Luus et al., 2015; Weng et al., 2017; Zhang et al., 2018a)[14-16],scene classification (Castelluccio et al., 2015)[17], change (Zhang et al.,2016) and object detection (Cheng et al., 2016; Zhang et al., 2018c)[18-19].Multi-layer artificial Convolutional Neural Network (CNN) allows au-tomatic extraction of high-level features from labeled images. By meansof convolutional kernels at multi-levels operating over upper-level feature maps, high-level features are extracted hierarchically through the network. The back-propagation strategy helps CNN adjust its net-work parameters automatically. The high generalization capacity of CNN outstands other machine learning algorithms and makes CNN themost mature and widely used deep learning framework

This research was supported by The People's Republic of China Ministry of Science and Technology [ 2018YFF0213606-03 (Mu Y., Hu T. L., Gong H., Li S. J. and Sun Y. H.) http://www.most.gov.cn], the Science and Technology Department of Jilin Province [20160623016TC, 20170204017NY, 20170204038NY,20200402006NC (Mu Y., Hu T. L., Gong H. and Li S.J.) http://kjt.jl.gov.cn], and the Science and Technology Bureau of Changchun City [18DY021 (Mu Y., Hu T. L., Gong H., and Sun Y. H.) http://kjj.changchun.gov.cn].

Mu Ye,Gong He,Hu Tianli,Li Shijun ,Sun Yu are with the College of Information Technology, Jilin Agricultural University,Changchun 130118, China, with the Jilin Province Intelligent Environmental Engineering Research Center,Changchun 130118, China,with the Jilin Province colleges and universities the 13th Five-Year Engineering Research Center,Changchun 130118, China,and also with the Jilin Province Agricultural Internet of Things Technology Collaborative Innovation Center, Changchun 130118,China(email:muye@jlau.edu.cn;gonghe@jlau.edu.cn;hutianli@jlau.edu.cn;lishijun@jlau.edu.cn; sunyu@jlau.edu.cn).

Ni Ruiwen, Zhang Chang, Zhang Tong are with the College of Information Technology, Jilin Agricultural University, Changchun 130118, China(e-mail:niruiwen@mails.jlau.edu.cn;zhangchang@mails.jlau.edu.cn; zhangtong@mails.jlau.edu.cn).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3090085, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing



(LeCun et al.,2015)[20].

Convolutional neural network has been widely used in the field of remote sensing. In 2019, in mineral exploration, it has been proposed: An Augmented Linear Mixing Model to Address Spectral Variability for Hyperspectral Unmixing[21]. In the aspect of hyperspectral, some suggestions have been put forward.For example:an iterative multitask regression framework for semi-supervised hyperspectral dimensionality reduction[22];When Non-Convex Modeling meets Hyperspectral Remote Sensing;graph convolutional networks for hyperspectral image classification[23];spatial-spectral manifold alignment for semisupervised hyperspectral dimensionality reduction[24].In the aspect of remote sensing image classification, some suggestions are put forward. For example: X-ModalNet: a Semi-Supervised deep Cross-Modal Network for classification of remote sensing data[25]; graph convolutional networks for Hyperspectral Image classification[26];more diverse means better: multimodal deep learning meets remote sensing imagery classification[27]；CoSpace: common subspace learning from hyperspectral-multispectral correspondences[28];a semi-supervised cross-modality learning framework for land cover and land use classification[29]；a Spatial-Frequency joint feature extractor for hyperspectral image classification[30]. There is also great progress in remote sensing positioning. In 2020, it has been proposed that Capture-aware identification of mobile RFID tags with unreliable channels[31]； a partitioning approach to RFID identification[32]； a time and energy saving based frame adjustment strategy (TES-FAS) tag identification algorithm for UHF RFID systems[33]； Idle slots skipped mechanism based tag identification algorithm with enhanced collision detection[34]； Redundant rule detection for software-defined networking[35].

However, objects must to be reshaped in different shapes and scales into the same size, because CNN requires a fixed size, This operation causes the loss of object shape and scale information. Furthermore, relationships of the surrounding neighbors (i.e. contextual information) should also be considered for accurate object classification. At the same time,there are still many problems: 1. Most CNN models are for high-precision and obvious features of remote sensing images; 2. The amount of parameters and memory of the model are large.In order to solve these problems, we improve the traditional VGG-16 model. The improved model has universal adaptability for high-precision and low-precision remote sensing images.And it is still applicable for remote sensing images with very low and fewer pixels. At the same time, the model also reduces the amount of parameters and the memory size of the model. Therefore, this paper proposes a lightweight model based on VGG-16 according to the above problems. Firstly, the network structure of the original VGG-16 model is optimized, and then the super parameters of the model are adjusted. The data in the data set is processed by our designed data processing method, and the processed data is input into the model for training.

## II. METHOD

### A. *Modeling*

VGG is a convolutional neural network model proposed that "very deep convolutional networks for large scale image recognition" by simonyan and zisserman in the document. The small convolution kernel is better than the large convolution kernel, because the multi-layer nonlinear layer can increase the network depth to ensure learning more complex patterns, and the cost is relatively small (less parameters).However, VGG consumes more computing resources and uses more parameters, resulting in more memory consumption.Most of the parameters come from the first fully connected layer, and VGG-16 has three fully connected layers.The images of the experimental data set belong to ultra-low pixel remote sensing images, which do not need multiple feature extraction and more parameters.Therefore, we will combine the original VGG-16 with the full convolution model, and reduce the parameters of the model and the number of layers of the full connection layer, which not only ensures the accuracy of the model feature extraction, but also realizes the lightweight of the model and improves the training speed of the model.

Our model is based on the original VGG-16 model and combined with the traditional full CNN model.First of all, we change the CNN layer to start from $32 \times 32$, add a max pooling layer after each convolution layer, with the size of $2 \times 2$, then add $64 \times 64$, $128 \times 128$, $256 \times 256$, $512 \times 512$ CNN layers, and add a max pooling layer ($2 \times 2$) after each CNN layer.We change the original $512 \times 512$ convolution that vgg-16 needs to go through twice to go through once, in which the size of the convolution layer is $3 \times 3$ and the step size is 2.

A nonlinear correction layer is added after each convolution layer, so that each layer uses two nonlinear correction layers instead of a single correction layer. Because the pixels of the dataset ,we use are very low and the feature points are not obvious enough.Adding two nonlinear correction layers can overcome the problem of gradient disappearance, reduce the phenomenon of over fitting, and improve the training speed of the model.

Our research mainly focuses on ultra-low pixel images, so we need to expand the size of the data set, otherwise the image itself may be small, after multi-layer convolution, it may not be able to output the image or reduce the accuracy of the model.In addition, there are few feature points in the low pixel image, so it is difficult to extract them accurately.To solve these problems, we add zero padding layer before each convolution layer, and choose $1 \times 1$ size. Zero padding layer is a method to increase the nonlinearity of decision function without affecting the acceptance domain of the volume layer. On the one hand, it starts to control the output of the volume layer, control the network structure.On the other hand, it can extract more detailed features. Therefore, it can be used to expand the data and extract the image features with low feature points accurately. Thus, the accuracy of the model we designed is further improved.

The model we designed changed the three layers of VGG-16 into two, which can effectively reduce the parameter of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3090085, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing


model, not reduce the accuracy of model recognition and classification, but also improve the training speed of the model. The first full connection layer has 4096 channels, the second performs 1000way ILSVRC classification, and the last layer is soft Max layer. In the network, the configuration of the full connection layer is the same.

After design and modification, our model results are shown in the following figure:

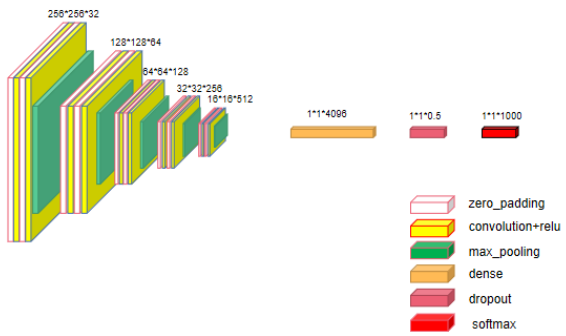


Fig 1. Model structure diagram. The figure shows the modified network model, showing the basic convolution layer and full connection layer.

## B. Image processing

Firstly, we transform the image size of the data set into 64 * 64, and then grayscale the image, and flatten the image into a data matrix（nxp:n=#sample ,p=#pixels in each image）.After that, the flat vector is inserted into the array, and the obvious features of the image are further displayed by string operation.Finally, divide the image pixels by 225 to scale the band.Now we take EuroSAT [36-37] data as an example.

The dataset is based on the collection of images taken by sentinel-2 satellite, covering 13 spectral bands, consisting of 10 classifications, and a total of 27000 land use images with labels and geographical references. It is used to detect the land use classification and land cover change and other issues to help improve the geographical environment. The data set includes the following 10 categories, each of which includes the following 10 categories, each containing 2000-3000 pictures, and the picture pixel is 64 × 64.

After image processing, we will process the image according to the ratio of 8:2 for the training set and test set. The following figure shows the data set processing process:

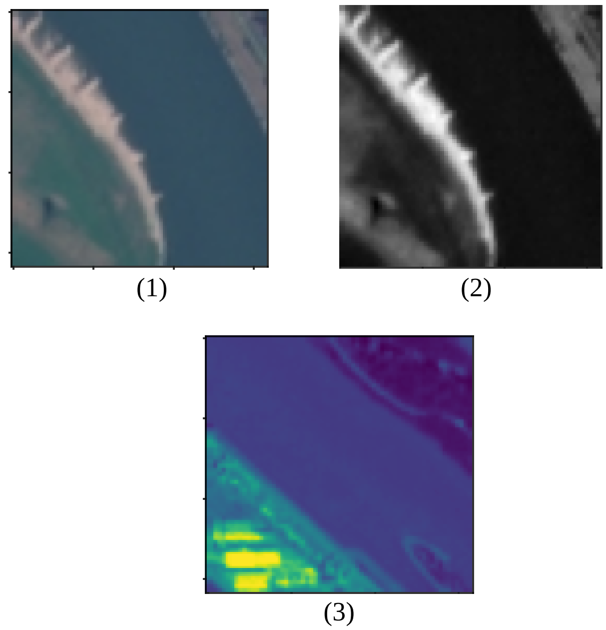


Fig 2. Data set processing.(1) is original image,(2) is graying,(3) is feature map.

## C. Training

In the model training, the batch size is set to 32, the momentum is set to 0.9, and the learning rate is 0.001.The training is regularized by weight attenuation and dropout regularization of two Denses (dropout is set to 0.5).Keras is used as the development framework in the experiment, and the average training time of each network model is 25 hours.The following figure shows the function diagram of the loss rate of training set and verification set of the model:

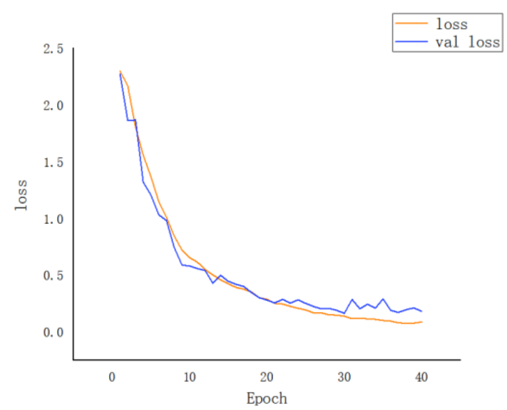


Fig 3. loss function diagram. The figure shows the function image of model training, the test loss is 0.2, and the val loss is 0.3.

## D. Testing

In the test, according to the following methods ,given a trained convolutional neural network and input images, convolutional neural networks are classified.First, it to its various orientations to the smallest predefined edge of the image.The network is then intensively applied to rescaled test images in a similar way. That is, the fully connected layer is converted to a convolution layer first. The resulting full convolutional network is then applied to the entire (untrimmed) image. The result is a class fraction map, the number of channels equal to the number of classes, and the variable spatial resolution depending on the size of the input image. Finally, in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3090085, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing



order to obtain a fixed-size vector of the remote sensing image of the image, the image is space-averaged (sum pool).

We input the test set into the model, and the function diagram of the accuracy of the test and verification set is as follows:

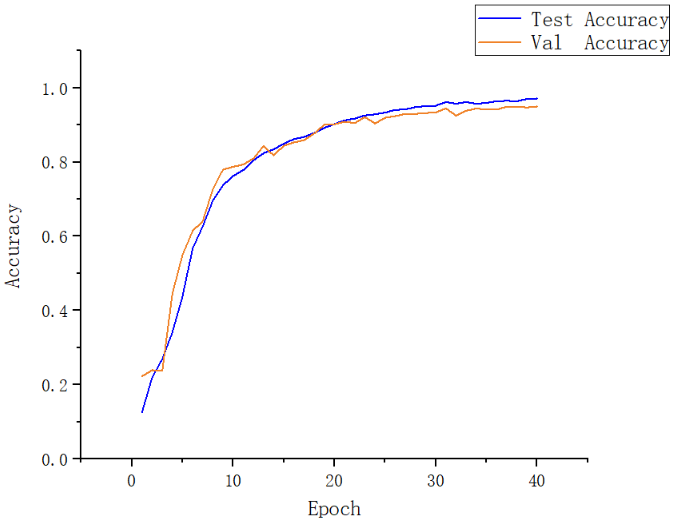


Fig 4. Accuracy function. The figure shows the function image of model testing, the test accuracy is 94%, and the val accuracy is 95%.

The loss rate function diagram of the accuracy function and the accuracy function diagram in this section have some fluctuations and overlaps. This is because the pixels of the image itself in the data set are too low and there are few obvious feature points, which leads to the image may be too similar. So the loss rate image and accuracy rate image are slightly abnormal. For example, forset and pastures, sealake and river.

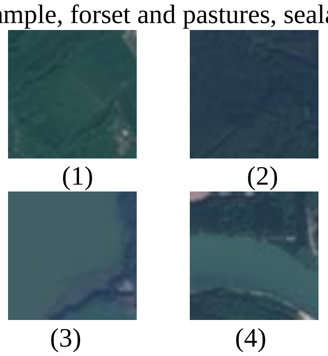


Fig 5.Image of similar data sets.(1) is the forest,(2) is the pastures,(3) is the sealake,(4) is the river.

## III. EXPERIMENTAL RESULTS AND ANALYSIS

### A. Experimental result

Our model is mainly aimed at remote sensing images with ultra-low pixels and less feature points, so we input EuroSAT(64 × 64) data set into the model, run it three times, and the average value of 10 classifications is the final accuracy, and the accuracy is 95%.

TABLE 1
OPERATING RESULTS OF EUROSAT IN OUR MODEL

| COUNT | SUM | RATE | LABEL |
|---|---|---|---|
| 482 | 451 | 0.9356 | RIVER |
| 612 | 564 | 0.9215 | ANNUALCROP |
| 616 | 550 | 0.8928 | HERBACEOUSVEGETATION |
| 503 | 480 | 0.9542 | INDUSTRIAL |
| 598 | 594 | 0.9933 | RESIDENTIAL |
| 528 | 478 | 0.9053 | HIGHWAY |
| 368 | 339 | 0.9211 | PASTURE |
| 597 | 594 | 0.9949 | FOREST |
| 591 | 585 | 0.9898 | SEALAKE |
| 505 | 449 | 0.8891 | PERMANENTCROP |

According to the above figure, it can be concluded that under the model, EuroSAT data classification can get a high accuracy. Our model simplifies convolution layer according to the remote sensing images of ultra-low pixels and few feature points, and adds zero padding layer to change the dimension size of the image and reduces the parameter quantity of model operation. In this way, the remote sensing images with low pixel and low feature points can be extracted more precisely; for high-precision and low-precision remote sensing images, we will use the more precise feature extraction small dimension is used to extract the image step by step, and the fine feature points of high-precision images are extracted, and the accuracy of recognition and classification is further improved.

As shown in Table 1, it can be seen that except permanentcrop and herbanceous vegetation, they all have high accuracy, especially forest, residential and sealake. Because the pixels of remote sensing images are very low, in the three data sets.The feature points are still obvious, which can be easily extracted for recognition, and the feature points of other types of images are also obvious, so we can also get higher accuracy.

It also can be seen that the accuracy of permanentcrop and herbaceous navigation is low.Because there are almost no feature points in the two datasets, and the images are mainly color differences. After image processing, the feature points are difficult to extract, so the accuracy of our model recognition is not high, In the future, we need to further improve the extraction of feature points.

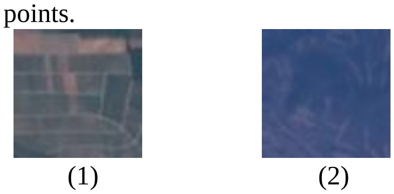


Fig 6. PermanentCrop and Herbanceous Vegetation.(1) is the PermanentCrop,(2) is the Herbanceous Vegetation.

### B. Analysis of experimental results

Analysis of experimental results Most of the current remote sensing data of the latest technology training model are 256 × 256 images. In order to see the applicability of our model for the remote sensing image with ultra-low pixels and less feature points more intuitively. We put the data set used in this experiment into the current latest model and run it three times. The best accuracy of the three times is the final accuracy, and the model runs accurately .The accuracy is shown in the figure:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3090085, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing



TABLE 2
OPERATNG RESULTS OF EUROSAT IN OTHER MODEL

| METHODS | 80% SAMPLES FOR TRAINING |
|---|---|
| OUR MODEL | 0.95 |
| VGG-16 | 0.79 |
| GOOGLENET | 0.85 |
| CAFFENET | 0.83 |
| RESNET50 | 0.90 |

According to table 2, we can see that the accuracy of our model for remote sensing image recognition and classification with very low pixels and few feature points is much higher than other models, while the accuracy of the model with high recognition accuracy in this dataset.It can not reach the effect of our model. This proves the superiority of our model in the recognition and classification of remote sensing images with very low pixels and few feature points.

At the same time, we input the data set selected by our model into the newly published (Attention Consistent Network for Remote Sensing Scene Classification) [38] The models and algorithms in classification are mainly for high-precision image classification, and because the pixels of the model we used are too low and the similarity is very high, there is over fitting phenomenon when the model runs EuroSAT dataset.

Table 2 and in the attention consistent network for remote sensing scene Experiments in classification show that: remote sensing image recognition is very different from natural image recognition, too complex and deep network is not suitable for ultra-low pixel remote sensing image recognition. Because deep convolution neural network has high semantic, when the diversity of training samples meet certain conditions. It is difficult to effectively improve the score by simply increasing the dimension of training data and features Class precision may lead to redundant information or noise.

In order to verify the universality of our model for high-precision and low-precision remote sensing images, experiments are carried out with representative remote sensing data sets.SIRI-WHO data (200 × 200)[39-41], OPTIMAL-31 (256 × 256) [42]and LEVIR (800 × 600)[43] are selected to test the model. There are many kinds of scenes in high-resolution remote sensing images, and there are great differences between the scenes without categories. For example, the scene analogy of aircraft and oil tank includes both the image of a single target and the image of multiple targets. After the traditional data set amplification method, the amplified images are processed according to the processing method in 2.2, and the processed images are input into our designed model for three experiments. The average value of the classification results is taken as the experimental result. The experimental results are shown in Table 3.

TABLE 3
OUR MODEL RUNS MODEL TEST RESULTS FROM OTHER DATASETS

| DATASET | COUNT | SUM | RATE |
|---|---|---|---|
| SIRI-WHO DATA | 1000 | 993 | 0.993 |
| OPTIMAL-31 | 1000 | 998 | 0.998 |
| WHO-RS19 | 1000 | 995 | 0.995 |

According to table 3, we can see that the accuracy of the experimental results is 99%, which shows that the nonlinear correction layer we added in the model can solve the problem of over fitting and gradient descent. At the same time, it also proves that although our model reduces the parameter amount, we can still guarantee the accuracy even though we change the convolution of 512 × 512 from twice to once, so that we can get the correct model The model memory is 2.3MB. At the same time, the pixels of the University of California dataset are 256 × 256, and the image feature points in the data are obvious, which is one of the reasons why the 99% accuracy can be obtained.

The results show that the scheme we designed is effective. The model designed by us is not only universal in high-precision and low-precision remote sensing images, but also in ultra-low-pixel and less feature points remote sensing images, which can reduce the requirements for the accuracy of remote sensing images in remote sensing image recognition.

## IV. CONCLUSION

In this paper, we modify the network model based on vgg-16, combine it with the traditional full convolution, fine tune the decoding part of the original network, and use the integrated learning strategy to optimize the prediction results of the model, so as to reduce the parameters of the model. Experiments show that our model has more than 98% accuracy in remote sensing image classification.

After comparing the recent models, we can find that our model is also suitable for remote sensing images with ultra-low pixel value or few feature points. Therefore, our model not only has good applicability in high-precision and low-precision remote sensing data subdivision, but also can classify fuzzy images and recognize and classify local features of images. Our model can reduce the requirements of image pixels in remote sensing image recognition and classification in the future, and further improve the efficiency of remote sensing image recognition and classification.

## REFERENCES

[1] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.
[2] G.-S. Xia et al., "AID: A benchmark data set for performance evaluation of aerial scene " *IEEE Trans. Geosci. Remote Sens. Lett*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.
[3] L. Fang, N. He, S. Li, P. Ghamisi, and J. A. Benediktsson, "Extinction profiles fusion for hyperspectral images classification," *IEEE Trans. Geosci. Remote Sens. Lett*, vol. 56, no. 3, pp. 1803–1815, Mar. 2018.
[4] G. Cheng, J. Han, L. Guo, Z. Liu, S. Bu, and J. Ren, "Effective and efficient midlevel visual elements-oriented land-use classification using VHR remote sensing images," *IEEE Trans. Geosci.Remote Sens. Lett*, vol. 53, no. 8, pp. 4238–4249, Aug. 2015.
[5] B. Zhao, Y. Zhong, G.-S. Xia, and L. Zhang, "Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens. Lett*, vol. 54, no. 4, pp. 2108–2123, Apr. 2016.
[6] X. Lu, X. Zheng, and Y. Yuan, "Remote sensing scene classification by unsupervised representation learning," *IEEE Trans. Geosci. Remote Sens. Lett*, vol. 55, no. 9, pp. 5148–5157, Sep. 2017.
[7] Q. Zhu, Y. Zhong, B. Zhao, G.-S. Xia, and L. Zhang, "Bag-of-visualwords scene classifier with local and global features for high spatial resolution remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 6, pp. 747–751, Jun. 2016.
[8] L. Huang, C. Chen, W. Li, and Q. Du, "Remote sensing image scene classification using multi-scale completed local binary patterns and Fisher vectors," *Remote Sens. Lett*, vol. 8, no. 6, p. 483, Jun. 2016.
[9] Y. Feng, Y. Yuan, and X. Lu, "Learning deep event models for crowd anomaly detection," *Neurocomputing*, vol. 219, pp. 548–556, Jan. 2017.

[10] X. Lu, B. Wang, X. Zheng, and X. Li, "Exploring models and data for remote sensing image caption generation," IEEE Trans. *Geosci. Remote Sens. Lett*, vol. 56, no. 4, pp. 2183–2195, Apr. 2018.

[11] W. Zhang, X. Lu, and X. Li, "A coarse-to-fine semi-supervised change detection for multispectral images," *IEEE Trans. Geosci. Remote Sens. Lett*, vol. 56, no. 6, pp. 3587–3599, Jun. 2018.

[12] W. Zhang, X. Lu, and X. Li, "A coarse-to-fine semi-supervised change detection for multispectral images," *IEEE Trans. Geosci. Remote Sens. Lett*, vol. 56, no. 6, pp. 3587–3599, Jun. 2018.

[13] J. Zhu, L. Fang, and P. Ghamisi, "Deformable convolutional neural networks for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, to be published, doi: 10.1109/LGRS.2018.2830403

[14] Luus.F.P.S.,Salmon,B.P.,Van Den Bergh,F.Maharaj,B.T.J., "Multiview deep learning for land-use classification," *IEEE Geosci. Remote Sens. Lett*,vol 12,pp 2448-2452,2015.

[15] Weng, Q., Mao, Z., Lin, J., Guo, W., "Land-use classification via extreme learning classifier based on deep convolutional features," *IEEE Geosci. Remote Sens. Lett*, vol.14,no.1,pp.704-708,2017.

[16] Zhang, Ce, Pan, X., Li, H., Gardiner, A., Sargent, I., Hare, J., Atkinson, P.M., "hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification," *ISPRS J. Photogramm. Remote Sens. Lett,* vol.104,no.10,pp.133-144,2018, 10.1016/j.isprsjprs.2017.07.014.

[17] Castelluccio, M., Poggi, G., Sansone, C., Verdoliva, L., "Land Use Classification in Remote Sensing Images by Convolutional Neural Networks," arXiv preprint arXiv:1508.0009,2015.

[18] Zhang, P., Gong, M., Su, L., Liu, J., Li, Z., "Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images." *ISPRS J. Photogramm. Remote Sens.*vol 116,no.10,pp.24-41,2016. 10.1016/j.isprsjprs.2016.02.013.

[19] Cheng, G., Zhou, P., Han, J., "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *Trans.Geosci.RemoteSens.*vol.54,no.2,pp,7405-7415,2016.

[20] Zhang, Chenxiao, Yue, P., Di, L., Wu, Z., "Automatic identification of center pivot irrigation systems from landsat images using convolutional neural networks," *Agriculture*,vol.8,no.10,pp.147,2018, 10.3390/agriculture8100147.

[21] Danfeng Hong, Naoto Yokoya, Jocelyn Chanussot, Xiao Xiang Zhu, "An Augmented Linear Mixing Model to Address Spectral Variability for Hyperspectral Unmixing," IEEE Transactions on Image Processing,vol 28,no 4,pp 1923-1938,2019.

[22] Danfeng Hong, Naoto Yokoya, Jocelyn Chanussot, Jian Xu, Xiao Xiang Zhu,"Learning to propagate labels on graphs: An iterative multitask regression framework for semi-supervised hyperspectral dimensionality reduction," ISPRS Journal of Photogrammetry and Remote Sensing, vol 158,pp.35-49,2019.

[23] Danfeng Hong, Wei He, Naoto Yokoya, Jing Yao, Lianru Gao, Liangpei Zhang, Jocelyn Chanussot, Xiao Xiang Zhu,"Interpretable Hyperspectral AI: When Non-Convex Modeling meets Hyperspectral Remote Sensing,"IEEE Geoscience and Remote Sensing Magazine,doi. 10.1109/MGRS.2021.3064051,2021.

[24] Danfeng Hong, Naoto Yokoya, Jocelyn Chanussot, Jian Xu, Xiao Xiang Zhu Joint and progressive subspace analysis (JPSA) with spatial-spectral manifold alignment for semisupervised hyperspectral dimensionality reduction,IEEE Transactions on Cybernetics, doi: 10.1109/TCYB.2020.3028931,2020.

[25] Danfeng Hong, Naoto Yokoya, Gui-Song Xia, Jocelyn Chanussot, Xiao Xiang Zhu. " X-ModalNet: A Semi-Supervised Deep Cross-Modal Network for Classification of Remote Sensing Data," ISPRS Journal of Photogrammetry and Remote Sensing, vol 167,pp. 12-23,2020.

[26] Danfeng Hong, Lianru Gao, Jing Yao, Bing Zhang, Antonio Plaza, Jocelyn Chanussot. "Graph Convolutional Networks for Hyperspectral Image Classification," IEEE Transactions on Geoscience and Remote Sensing, doi : 10.1109/TGRS.2020.3015157,2020.

[27] Danfeng Hong, Lianru Gao, Naoto Yokoya, Jing Yao, Jocelyn Chanussot, Qian Du, Bing Zhang. "More Diverse Means Better: Multimodal Deep Learning Meets Remote Sensing Imagery Classification, IEEE Transactions on Geoscience and Remote Sensing," vol. 59,no. 5,pp.4340-4354,2021.

[28] Danfeng Hong, Naoto Yokoya, Jocelyn Chanussot, Xiao Xiang Zhu, "CoSpace: Common Subspace Learning from Hyperspectral-Multispectral Correspondences," IEEE Transactions on Geoscience and Remote Sensing, vol. 57,no. 7,pp.4349-4359,2019.

[29] Danfeng Hong, Naoto Yokoya, Nan Ge, Jocelyn Chanussot, Xiao Xiang Zhu, "Learnable Manifold Alignment (LeMA) : A Semi-supervised Cross-modality Learning Framework for Land Cover and Land Use Classification," ISPRS Journal of Photogrammetry and Remote Sensing,vol. 147,pp. 193-205,2019.

[30] Danfeng Hong, Xin Wu, Pedram Ghamisi, Jocelyn Chanussot, Naoto Yokoya, Xiaoxiang Zhu. "Invariant Attribute Profiles: A Spatial-Frequency Joint Feature Extractor for Hyperspectral Image Classification,"IEEE Transactions on Geoscience and Remote Sensing,vol. 58,no, 6,pp. 3791-3808,2020.

[31] J. Su, Z. Sheng, A. X. Liu, Y. Han, and Y. Chen, "Capture-aware identification of mobile RFID tags with unreliable channels,"IEEE Transactions on Mobile Computing, pp. 1-14, doi: 10.1109/TMC.2020.3024076,2020.

[32] J. Su, R. Xu, S. Yu, B. Wang, and J. Wang, "Idle slots skipped mechanism based tag identification algorithm with enhanced collision detection," KSII Transactions on Internet and Information Systems, vol. 14, no. 5, pp. 2294-2309, 2020.

[33] J. Su, Z. Sheng, A. Liu, Z. Fu, and Y. Chen, " A time and energy saving based frame adjustment strategy (TES-FAS) tag identification algorithm for UHF RFID systems," IEEE Transactions on Wireless Communications, vol. 19, no. 5, pp. 2974-2986, 2020.

[34] J. Su, R. Xu, S. Yu, B. Wang, and J. Wang, "Redundant rule detection for software-defined networking," KSII Transactions on Internet and Information Systems, vol. 14, no. 6, pp. 2735-2751, 2020.

[35] J. Su, Z. Sheng, A. Liu, Z. Fu, and Y. Chen, " A time and energy saving based frame adjustment strategy (TES-FAS) tag identification algorithm for UHF RFID systems," IEEE Transactions on Wireless Communications, vol. 19, no. 5, pp. 2974-2986, 2020.

[36] Helber, Patrick and Bischke, Benjamin and Dengel, Andreas and Borth, Damian, "A novel dataset and deep learning benchmark for land use and land cover classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,vol.12,no.7,pp.2217-2226,2019.

[37] Helber, Patrick and Bischke, Benjamin and Dengel, Andreas and Borth, Damian, "A novel dataset and deep learning benchmark for land use and land cover classification," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium,pp.201-207,doi: 10.1109/IGARSS.2018.8519248.

[38] X Tang, QS Ma, XR Zhang, F Liu,JJ Ma,LC Jiao, "Attention Consistent Network for Remote Sensing Scene Classification, " *IEEE journal of selected topics in applied earth observations and remote sensing*, vol 14,pp. 2030-2045,2021.

[39] B. Zhao, Y. Zhong, G.-s. Xia, and L. Zhang, "Dirichlet-Derived Multiple Topic Scene Classification Model Fusing Heterogeneous Features for High Spatial Resolution Remote Sensing Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 4, pp. 2108-2123, 2016.

[40] B. Zhao, Y. Zhong, L. Zhang, and B. Huang, "The Fisher Kernel Coding Framework for High Spatial Resolution Scene Classification," *Remote Sensing*, vol. 8, no. 2, p. 157, doi:10.3390/rs8020157 2016.

[41] Q. Zhu, Y. Zhong, B. Zhao, G.-S. Xia, and L. Zhang, "Bag-of-Visual-Words Scene Classifier with Local and Global Features for High Spatial Resolution Remote Sensing Imagery," *IEEE Geoscience and Remote Sensing Letters*, doi:10.1109/LGRS.2015.2513443,2016.

[42] Q. Wang, M. Chen, F. Nie, and X. Li, "Detecting Coherent Groups in Crowd Scenes by Multiview Clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, vol. 42, no. 1, pp. 46-58, 2020.

[43] Bin Pan, Zhenwei Shi and Xia Xu, "Hierarchical Guidance Filtering Based Ensemble Classification for Hyperspectral Image, ",*IEEE Transactions on Geoscience and Remote Sensing*,vol. 55,no. 7,pp. 4177-4189,2017.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3090085, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing



**Mu Ye** received his Bachelor of Science degree from the School of Electronic Science and Engineering of Jilin University in China in 2003 and his Ph.D. in Science from the School of Electronic Science and Engineering of Jilin University in China in 2016

He is currently a lecturer at Jilin Agricultural University in China, focusing on bio health function-environmental model research and artificial intelligence and intelligent agriculture research.

**Ni Ruiwen** received his bachelor's degree in computer science and technology from Jilin Agricultural University in China in 2019 and is a graduate student in computer science and technology at Jilin Agricultural University.

Her research interests are artificial intelligence and intelligent agriculture.

**Zhang Chang** received a bachelor's degree in electronic information science and technology from Jilin Agricultural University in 2019. Currently studying for a graduate degree, the main research direction is machine learning.

**Gong He** received a bachelor's degree in electrical information engineering from Changchun University of Technology in China in 2002 and a master's degree in electrical and communication engineering from the School of Communications of Jilin University in China in 2010.

He is currently an associate professor at Jilin Agricultural University, whose main research interests are agricultural Internet of Things technology, agricultural engineering intelligence, embedded technology.

**Li Shijun** received his Bachelor of Science degree in Physics from Shiping Normal College in China in 1991.

He is currently a professor at Jilin Agricultural University in China, focusing on Internet of Things engineering, architecture research an application.

**Hu Tianli** received his Bachelor of Science degree from Jilin University School of Physics in 2008 and his Ph.D. in Science from Jilin University in China in 2013

He is currently a lecturer at Jilin Agricu ltural University in China, focusing on facility agricultural environmental mon itoring and intervention, sensor networ king, artificial intelligence and digital a griculture.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3090085, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing



**Sun Yu** received his bachelor's degree in electronic information engineering from Changchun University of technology in 2009 and his master's degree in communication engineering from Sydney University of science and technology in 2014.

He is now a teaching assistant in Jilin Agricultural University of China. His main research directions are fungi, bioinformatics, artificial intelligence and communication network.

**Zhang Tong** received the B.S. degree in Software Engineering,Quancheng College, University of Jinan,China,in 2019 and currently studying Agricultural Engineering and Information Technology at Jilin Agricultural University. She is majors in artificial intelligence.

**Guo Ying** obtained a master's degree in electronic and communication engineering from Jilin University of China in 2013. At present, she is a lecturer in Jilin Agricultural University of China. Her main research direction is artificial intelligence and intelligent agriculture.