

Early and late rice identification from Tiangong-2 wide band images based on CNN

Zhuang Zhou^{1,2*}, Shengyang Li^{1,2}, and Yuyang Shao^{1,2}

¹Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, China

²Key Laboratory of Space Utilization, Chinese Academy of Sciences, China

Abstract. The wide band images acquired from the Tiangong-2 space laboratory covers many spectral bands such as visible light, shortwave infrared and thermal infrared. These high-quality images can be used for space science experiments such as earth observation. In this paper, we use CNN (convolutional neural networks) to extract the spectral features of different landcover from the wide band images, then identify the early rice and the late rice accurately in Huarong County, Hunan Province, China. With advanced techniques such as deep learning, the spatial distribution information of crops can be effectively obtained from the wide band images which can provide data services for agricultural production management.

1 Introduction

Tiangong-2 space laboratory was successfully launched on 15th September 2016. After the successful mission of Tiangong-1 space laboratory, Tiangong-2 is the second space laboratory emission launched by China. Tiangong-2 has carried out a number of new space applied load equipment and more than ten applications and experiments, which relate to earth observation and space science. The wide band imaging spectrometer is one of the payload in Tiangong-2 which has wide FOV (Field of View) and combination of image and spectrum.

It is the first sensor which integrates visible and near infrared, short wavelength infrared and thermal infrared spectrums in one instrument in the world. It is a push-broom sensor that is channel programmable in the visible near infrared spectrums and multi-spectral detection in short wavelength and thermal infrared spectrum. There are 14 band channels with a spatial resolution of 100 meters which is mainly aimed at medium resolution and large-scale object monitoring. It is suitable for the study of landcover classification, as well as for ocean and coastal zone water color and water temperature observation.

Numerous traditional landcover classification methods have been applied in the field of earth observation, such as k-nearest-neighbors, minimum distance, support vector machines (SVM), random forest and logistic regression [1-3]. Recently, some more effective feature extraction methods as well as advanced classifiers were proposed, such as deep learning. In the past few years' deep learning are the most popular and efficient approaches for

* Corresponding author: zhouzhuang@csu.ac.cn

landcover classification of remote sensing [4]. More and more application results show that the methods based on deep learning are superior to SVM. As a powerful machine learning method, deep learning can be applied in solving various tasks in the fields of image classification, speech recognition and natural language processing. The main idea of deep learning is to simulate the human visual system and extract effective features from the information in a hierarchical manner. In the field of image classification, a large number of deep learning models, frameworks and standard data sets are available for reference [5].

Convolutional neural networks (CNN) is one of the most popular models of deep learning, which is widely used in dealing with image vision problems. The weight-sharing structure of CNN makes it similar to the biological neural network which can reduce the complexity of the model and has prominent advantages in two-dimensional image processing. The idea of CNN was first proposed in [6], further developed in [7], and simplified and refined in [8]. CNN has been demonstrated to provide excellent classification performance than traditional classifiers in earth observation area. However, the research of remote sensing classification based on deep learning mainly for high-resolution or hyperspectral remote sensing images which are difficult to obtain. To explore the feasibility of crops classification and recognition from wide band images based on deep learning is of great significance.

2 The study area and data

2.1 The study area

The purpose of this paper is to achieve the landcover and crops classification from Tiangong-2 wide band images for Huarong county, which located in the northeast of Hunan province, China, with an area of around 1642 square km (112.31 ~113.03°E; 29.17~29.81°N). The climate of the study area is subtropical humid with an annual rainfall of 1304mm and average temperature of 17.2 °C. The superior geographical location and natural conditions make Huarong County an important rice producing area.

The crops planting structure of this region is stability which are mainly composed of early rice and late rice from May to October every year. Early rice is transplanted in mid-April and harvested in July, late rice is transplanted in July and harvested in October. The main landcover of the study area is classified into five types such as early rice, late rice, forest, buildings and water. The false color composite image of the study area from Tiangong-2 is shown in figure 1.

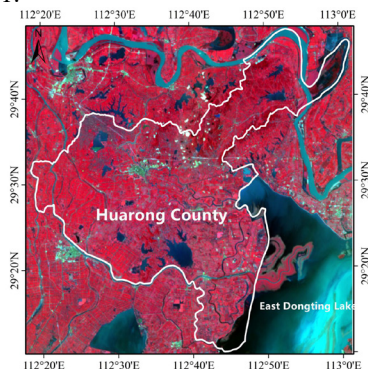


Fig. 1. The false color composite image of study area from Tiangong-2.

2.2 Data

The wide band images from Tiangong-2 we selected were acquired in July 25, 2017, which contains 14 bands of channels and has a spatial resolution of 100 meters. Abundant spectral information provides effective data source for rice feature extraction and identification. In order to obtain accurate sample label data, we selected the Sentinel-2A high-resolution remote sensing images (transiting time of July 17, 2017, spatial resolution of 10m) in similar period for visual interpretation and collected the main five types of ground object samples. Through the spatial position matching, the corresponding spectral vectors of the samples in the wide band images of Tiangong-2 are extracted, resampled and transformed into label images. A total 179 different types of polygons were collected by artificial visual interpretation which divided into training and testing data sets to train and validate the classifier, respectively. All of the labeled pixel data we collected were randomly divided into training set and verification set according to the proportion of 1:1.

3 The research method

3.1 CNN

CNN is a feed-forward neural network which composed of input layer, convolutional layer, pooling layer, fully connected layer and output layer. With local connections and weights sharing, the number of parameters can be drastically reduced. In addition, the risk of over-fitting can also be decreased. A normal CNN network usually consists of several pairs of convolutional layer, max pooling layer and ends with a full connection layer. The architecture of CNN classifier constructed in this paper is shown in figure 2.

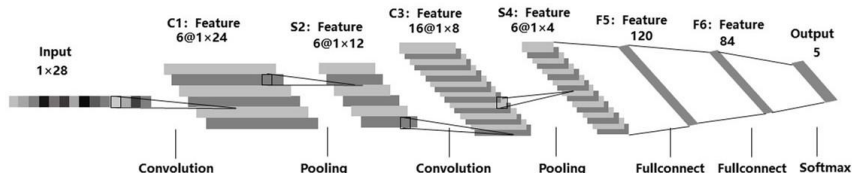


Fig. 2. The architecture of the proposed CNN classifier.

In a regular neural network, every neuron was connected to all neurons in the next layer. In contrast, the neurons on the convolutional layer of CNN are sparsely connected to the neurons in the max pooling layer based on their relative location. Neurons that belong to the same layer share the same weights which reduce the number of model parameters and make the model training process more efficient. The max pooling layer is generally attached behind the convolutional layer.

Max pooling partitions the input data into a set of nonoverlapping windows which can select the maximum value for each sub region and reduces the calculation complexity of model calculations and enhance the invariance of feature translation. In order to classify, the computation chain of the CNN ends with a fully connected network, which integrates all the feature maps of different locations.

3.2 CNN-based classification

The hierarchical structure of CNN has proved to be the most effective way to learn image representation in visual field. Spectral vectors of 14 bands from the wide band images are resampled and converted to grayscale images with dimension size 28 in figure 3.

As illustrated in figure 2, the CNN network constructed in this paper consists of two convolutional layers, each of which is connected to a max pooling layer. The last two layers of the model are fully connected layers. The rectified linear unit (ReLU) function was

chosen as the activation function which is also the most popular and effective activation function in deep learning field. Compared to the sigmoid function, the advantages of using ReLU enables more efficient computations and faster gradient diffusion for training CNN.

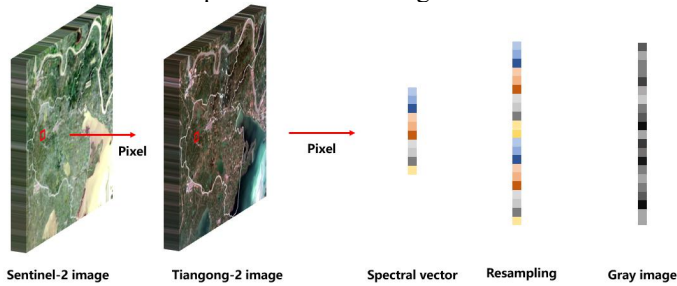


Fig. 3. The gray image processing.

The difference of CNN network structure mainly lies in the number of network layers and the training mode of network. As shown in Figure 2, the network consists of eight different layers, including the input layer, two convolution layers, two maximum pool layers, two full connection layers and an output layer.

In wide band images, each pixel sample can be viewed as a two-dimensional gray image with a height of 1. Therefore, the size of the input layer is just $\lceil 1 \times 28 \rceil$, and 28 is the resampled to 2 times number of bands. The $\lceil 1 \times 28 \rceil$ input data was filtered by the first hidden convolutional layer with 6 kernels of size $\lceil 1 \times 5 \rceil$.

The first hidden convolutional layer contains $\lceil 6 \times 1 \times 24 \rceil$ nodes, and $\lceil 24 = 28 - 5 + 1 \rceil$. The number of trainable parameters between the first two layers are $\lceil 6 \times (5 + 1) \rceil$. The first max pooling layer is the second hidden layer, and the kernel size is $\lceil 1 \times 2 \rceil$. This layer contains $\lceil 6 \times 1 \times 12 \rceil$ nodes, and $\lceil 12 = 24 / 2 \rceil$.

The second convolutional layer contains $\lceil 16 \times 1 \times 8 \rceil$ nodes, and $\lceil 8 = 12 - 5 + 1 \rceil$. The kernel size of the second max pooling layer is $\lceil 1 \times 2 \rceil$. This layer contains $\lceil 6 \times 1 \times 4 \rceil$ nodes, and $\lceil 4 = 8 / 2 \rceil$. All the fully connected layers connected with the softmax classifier where output number is 5.

According to the corresponding CNN and the afore mentioned parameters, classifying a specified pixel of wide band images can be realized. In our architecture, convolutional layers and max pooling layers can be viewed as a trainable feature extractor to the input spectral information of wide band images, and the second fully connected layer is trainable classifier to the feature extractor. The output of sampled is the actual characteristics of raw data. The CNN model designed in this paper can finally extract 120 kinds of features from original images.

All the trainable parameters in our CNN should be initialized to be a random value between -0.05 and 0.05. The first step of training process is forward propagation which aims to compute the actual classification result of the input data with current parameters. The second training process step is back propagation which is employed to update the trainable parameters in order to make the discrepancy between the actual classification output and the desired classification output as small as possible.

Because the structure and all the relevant weight values are specified, we can build the CNN classifier and reload the saved parameters classifying wide band images, the classification process is just like the forward propagation step.

4 Experiment

All the programs are implemented using Caffe which is a CNN framework that makes us easily define, optimize, and evaluate mathematical expressions involving multidimensional arrays efficiently and conveniently on GPUs. The results are generated on a PC equipped with an Intel Core i7 with 2.6 GHz and Nvidia GeForce GTX 530 graphics card.

4.1 Date preprocessing

The main five types of ground object samples were collected from the Sentinel-2A high-resolution images by visual interpretation. Then, through the spatial position matching, the corresponding spectral vectors of the samples in the wide band images of Tiangong-2 are extracted, resampled and transformed into label images. Available training data are divided into training and testing samples for training the parameters and calculating the model loss of the proposed CNN classifier. Moreover, each pixel is scaled to [0,1] uniformly.

4.2 Results and comparisons

Figure 4 illustrate the distributions of early rice and late rice of Huarong County in 2017 from Tiangong-2 wide band images by CNN. The results showed that the early rice mainly concentrated in the west of Huarong County, showing a continuous and massive distribution, indicating that the early maturing rice in the western region had not yet been harvested. The distribution of early rice and late rice in the eastern and central regions was rather fragmented, reflecting that the harvesting and transplanting alternating features.

To demonstrate the relationship between accuracies and the differences species of landcover, we present the detail accuracies of our proposed CNN classifiers in Table 1. The overall accuracy of corresponding classification is 95.96% and kappa coefficient is 0.95. The accuracies more than 90% for all classes of major agricultural crops (early rice and late rice).

The excellent performance verifies that the proposed CNN classifier has discriminative capability to extract subtle visual features, which is even superior to human vision for classifying complex curve shapes.

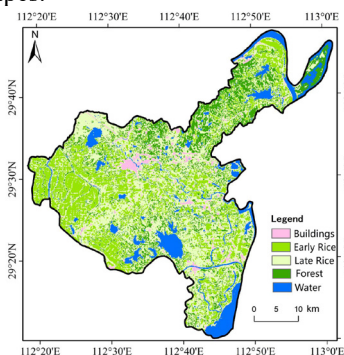


Fig. 4. The spatial distributions of the peanut in Huarong.

Table 1. The classification accuracy of Huarong.

	Buildings	Water	Forest	Early Rice	Late Rice	Total	UA
Buildings	97	0	0	2	2	101	96.04%
Water	0	286	0	0	0	286	100.00%
Forest	0	0	87	1	1	89	97.75%
Early Rice	0	6	0	173	4	183	94.54%

Late Rice	0	0	9	5	150	164	91.46%
Total	97	292	96	181	157	823	
PA	100.00%	97.95%	90.63%	95.58%	95.54%		
OA	95.96%	Kappa	0.95				

5 Conclusion

In this paper, we proposed a CNN-based method for early and late rice planting area extract on wide band images from Tiangong-2. Our work is an exploration of using CNN for crops recognition and achieves excellent performance. The architecture of our proposed CNN classifier contains two convolutional layers, two max pooling layers and two fully connected layers, due to the small number of training samples. In the feature, some techniques, such as Dropout, can be used to alleviate the overfitting problem caused by limited training samples. Furthermore, only spectral information of crops was considered in our work, we should pay more attention to multi-features such as time phases information of crops, to improve the result of classification.

In addition, recent researches in deep learning have indicated that unsupervised learning can be employed to train CNN, reducing the requirement of labeled samples significantly. Deep learning, especially deep CNN, should have great potentiality for remote sensing classification in the future. Moreover, in the current work, we do not consider the spatial correlation and only concentrate on the spectral signatures. We believe that some spatial-spectral techniques also can be applied to further improve the CNN based classification.

References

1. Brown J C, Kastens J H, Coutinho A C, et al. Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data. *Remote Sensing of Environment*, 2013, 130: 39-50.
2. Li X, Xu X, Wang J, et al. Crop classification recognition based on time-series images from HJ satellite. *Transactions of the Chinese Society of Agricultural Engineering*, 2013, 29(2): 169-176. (in Chinese with English Abstract)
3. Khatami, Reza, Giorgos Mountrakis, and Stephen V. Stehman. "A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research." *Remote Sensing of Environment* 177 (2016): 89-100.
4. Chen, Yushi, et al. "Deep learning-based classification of hyperspectral data." *IEEE Journal of Selected topics in applied earth observations and remote sensing* 7.6 (2014): 2094-2107.5.
5. Kussul, Nataliia, et al. "Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data." *IEEE Geoscience and Remote Sensing Letters* 14.5 (2017): 778-782.
6. Fukushima K. Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural networks*, 1988, 1(2): 119-130.
7. LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
8. Simard P Y, Steinkraus D, Platt J C. Best practices for convolutional neural networks applied to visual document analysis. //ICDAR. 2003, 3: 958-962.