Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering)
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.jzus.zju.edu.cn; www.springerlink.com
E-mail: jzus@zju.edu.cn

Classification of hyperspectral images based on a convolutional neural network and spectral sensitivity*

Cheng-ming YE†1,2, Xin LIU†‡2, Hong XU1,3, Shi-cong REN1,3, Yao LI4, Jonathan LI5
1*Chongqing Engineering Research Center of Automatic Monitoring for Geological Hazards, Chongqing 401120, China*
2*Key Laboratory of Earth Exploration and Information Technology of Ministry of Education,*
Chengdu University of Technology, Chengdu 610059, China
3*National Breeding Base of Technology and Innovation Platform for Automatic-monitoring of Geologic Hazards,*
Chongqing Institute of Geology and Mineral Resources, Chongqing 401120, China
4*Key Laboratory of Mountain Hazards and Earth Surface Process, Chinese Academy of Sciences, Chengdu 610041, China*
5*Department of Geography and Environmental Management, University of Waterloo, Waterloo, N2L 3G1, Canada*
†E-mail: rsgis@sina.com; astluxn@outlook.com

Received Mar. 7, 2019; Revision accepted Oct. 9, 2019; Crosschecked Mar. 4, 2020

Abstract: In recent years, deep learning methods have gradually come to be used in hyperspectral imaging domains. Because of the peculiarity of hyperspectral imaging, a mass of information is contained in the spectral dimensions of hyperspectral images. Also, different objects on a land surface are sensitive to different ranges of wavelength. To achieve higher accuracy in classification, we propose a structure that combines spectral sensitivity with a convolutional neural network by adding spectral weights derived from predicted outcomes before the final classification layer. First, samples are divided into visible light and infrared, with a portion of the samples fed into networks during training. Then, two key parameters, unrecognized rate (δ) and wrongly recognized rate (γ), are calculated from the predicted outcome of the whole scene. Next, the spectral weight, derived from these two parameters, is calculated. Finally, the spectral weight is added and an improved structure is constructed. The improved structure not only combines the features in spatial and spectral dimensions, but also gives spectral sensitivity a primary status. Compared with inputs from the whole spectrum, the improved structure attains a nearly 2% higher prediction accuracy. When applied to public data sets, compared with the whole spectrum, on the average we achieve approximately 1% higher accuracy.

Key words: Hyperspectral imaging; Deep learning; Convolutional neural network (CNN); Spectral sensitivity
https://doi.org/10.1631/jzus.A1900085 **CLC number:** TP751

1 Introduction

Using hyperspectral sensors, a continuous electromagnetic spectrum of each pixel in a scene is obtained. Hyperspectral images have been used in many fields (Yang et al., 2006; Bioucas-Dias et al., 2013; Wentz et al., 2014; Ye et al., 2017; Onojeghuo et al., 2018). Extraction of information and features hidden in hyperspectral images are the preconditions for classification, identification, and segmentation. With advances in the development of equipment, it is now possible to acquire a larger number of images, and this requires more efficient methods of processing.

The idea of the neural network (NN) originated

---

‡ Corresponding author

* Project supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA23090203), the National Key Technologies Research and Development Program of China (No. 2016YFB0502600), and the Key Program of Sichuan Bureau of Science and Technology (No. 2018SZ0350), China

ORCID: Cheng-ming YE, https://orcid.org/0000-0002-6799-0286; Xin LIU, https://orcid.org/0000-0002-0409-7466

in 1943 (McCulloch and Pitts, 1943) and was introduced in machine learning (ML) in 1958 (Rosenblatt, 1958). However, it was not until 2006, when Hinton and Salakhutdinov (2006) solved the gradient vanishing problem, that deep learning (DL) became prevalent. It was AlexNet (Krizhevsky et al., 2017) that convinced deep learning communities that DL really works in computer vision. Thus, the convolutional neural network (CNN) became the most popular algorithm in the computer vision field and was applied gradually to other fields (Cai et al., 2015; Wang et al., 2019).

Extracting features from remote sensing images is also a way to process images. Therefore, it is reasonable to apply the ML algorithms to hyperspectral imaging (HSI), which helps in classification and identification.

To more efficiently classify land-cover types in hyperspectral images into selected classes, some supervised ML algorithms have been applied to the hyperspectral remote sensing domain (Du and Zhang, 2014; Sigurdsson et al., 2014; Zhang et al., 2016; Gu et al., 2017). However, the lack of labeled samples, as well as the high dimension of hyperspectral data, has led to the Hughes phenomenon (Li et al., 2010). Support vector machines (SVM) have proved a good solution to this problem (Melgani and Bruzzone, 2004), but they are too simple to extract deep features in data sets (Bengio et al., 2013).

DL methods, proposed in recent research for classification and identification, not only avoid the Hughes phenomenon, but also provide a much higher accuracy of image classification and identification.

Compared with other ML algorithms, DL methods extract both low-level and high-level features. By combining spectral, textural, and semantic features in decision layers, classification results improve considerably. Cao et al. (2016) proposed a structure with a pixel-wise classifier and conducted superpixel segmentation on a prediction map, jointly predicting images by spatial and spectral features. Their experimental results show that this method is more competitive than common CNNs. Zhong et al. (2018) proposed an end-to-end spectral-spatial residual network (SSRN) based on ResNet. To enhance spatial features, Li et al. (2017) presented a reconstruction model based on deep convolutional neural networks (DCNNs). To process high-level features, Zhao and Du (2016) proposed a multiscale convolutional neural network (MCNN), which, for urban areas, is more accurate than common CNNs. The deep belief network (DBN), another popular deep model, has also been used in the HSI domain. By combining the extraction and classification of spectral-spatial features (Chen et al., 2015), DBN achieves an average of about 1% higher prediction accuracy for some public data sets. During pretraining and training, a diversified DBN was developed to obtain higher accuracies (an average of nearly 1%) and deal with the deficiency in training samples (Zhong et al., 2017). Additionally, the recurrent neural network (RNN) analyzes and processes hyperspectral images as sequential data (Mou et al., 2017), and also performs well on some airborne hyperspectral images.

In remote sensing images, different land-cover types have different spectral sensitivities. Some objects are sensitive to visible light; whereas, others are sensitive to infrared. Differentiating the contribution of various objects to different bands helps to more accurately and efficiently classify the land-cover types.

We propose a classification algorithm based on CNN and spectral sensitivity. According to the wavelength of each band, we divide origin images and labeled samples into the visible and the infrared spectral regions. Using their corresponding samples, we train these two parts, then predict and output two classification results. After comparing the two classification results, we select six segments where the accuracies of land-cover types, one each in visible and infrared outputs, exhibit the greatest discrepancy between each other. For these land-cover types, spectral weights, derived from accuracies, are added to the classification layer. This helps to improve overall accuracy. Finally, we compare the new algorithm with the original outputs. To prove the generalization of the proposed method, we apply this method to two public data sets: the Kennedy Space Center (KSC) data set and the Botswana data set.

2 Data and method

2.1 Data

Data was obtained in Beijing, China, onboard an aircraft, by a hyperspectral sensor with 64 continuous bands, ranging from 462 nm to 10 250 nm. The first 27 bands are within the visible light region; the remainder are in the infrared. Besides high spectrum resolution, the hyperspectral sensor also has high spatial resolution of 1 m per pixel, which is convenient for selecting accurate samples. The image has 3629 rows and 513 lines (1 861 677 pixels in total).

The KSC data set was obtained at the Kennedy Space Center, Florida, USA by the NASA AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) airborne sensor, with 224 continuous bands ranging from 400 nm to 2500 nm. Because of the influence of noise, only 176 spectral bands were retained. The KSC data set was labeled into 13 different land-cover types. The Botswana data set was acquired over the Okavango Delta, Botswana by the NASA EO-1 satellite. The Botswana data set, labeled into 14 classes, has the same wavelength range as the KSC data set, but with 242 spectral bands, 145 of which are retained.

2.2 Samples and input

Samples, including buildings, roadways, concrete ground, arable land, vegetation, and water, were selected and labeled (Fig. 1).

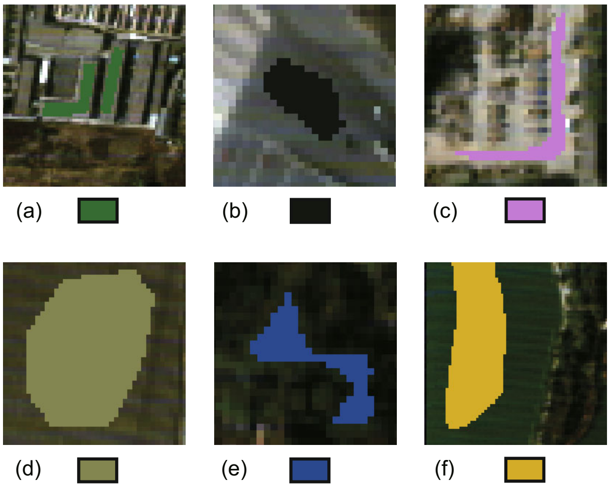

Fig. 1 Parts of selected samples shown on GIS software: (a) building; (b) roadways; (c) concrete groud; (d) arable land; (e) vegetation; (f) water

Selected as samples were 6720 pixels, belonging to six different land-cover types. A labeled pixel and its neighborhoods, with the same label as the central pixel, were constructed as a training sample for CNN.

The label of a resultant picture represents not all pixels, but only the central pixels in the picture, which means the central pixels in the hyperspectral image are formed as input pictures for CNN. Table 1 shows, in detail, the number of different types.

Table 1 Numbers of different samples, split into training and test samples

Sample name	Number of samples		Total
	Training	Test	
Building	719	720	1439
Roadway	167	167	334
Ground	63	64	127
Arable land	659	660	1319
Vegetation	47	48	95
Water	703	703	1406
Total	2358	2362	6720

In addition to the training/test proportion (Table 1), we also tried other different proportions such as 5% and 10%. Different from our data set, we took 20% of all pixels in public data sets as training sets, 30% as validation sets, and 50% as test sets.

Different structures of networks have different inputs. To independently train samples in both visible and infrared bands and compare their outcomes with ordinary structures, we constructed three sets of inputs. Regarding spatial dimensions, input pictures were formed on a scale of 5×5. Considering the high redundancy in the spectral dimension, principal component analysis (PCA) was used to convert correlated spectral data into linearly uncorrelated data (Hotellings, 1933), which reduced the complexity of the whole network and made it easier for the network to converge. In this process, we conserved more than 99% of the variance, which corresponds to five principal components in the spectral dimension. Finally, the scale of the input pictures of an ordinary structure was established as a $5 \times 5 \times 5$ 3D structure.

2.3 Convolutional neural network

A conventional CNN has four main layers: input, convolutional, full-connected (FC), and classification. Two main procedures, training and predicting, are included.

As a supervised algorithm, labeled pictures were fed into input layers as training samples and then convoluted by kernels in convolutional layers. After a series of convoluting and pooling, deep features were extracted and flattened into a column of neurons, which serves as the input layer for subsequent

FC layers. FC layers connect features and corresponding labels by a set of polynomials and send errors backward to adjust weights and bias to meet their correlations. Training extracts features and excavates correlations between features and corresponding labels.

Then prediction follows. By feeding pictures to be classified, a well-tuned CNN extracts their features and classifies them according to the polynomials tuned at training.

By sliding onto the output of the former layer, a convolution kernel (in other words, a weight matrix) extracts features from the picture and forms a feature map (LeCun et al., 1989), which is convoluted by kernels in the latter layers. This process is shown in Fig. 2.

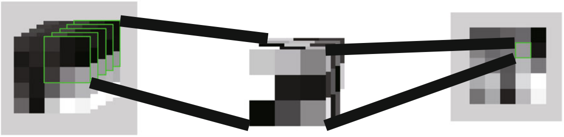

Fig. 2 Diagram of convolution

The outcome of convolution is fed into a non-linear activation function, rectified linear unit (ReLU), which allows a network to obtain easily sparse representation (Glorot et al., 2011) and avoids the vanishing gradient problem. Given a convolution result, $\boldsymbol{z}$, ReLU activates it as follows:

$$\phi(z_{ij}) = \max(0, z_{ij}), \tag{1}$$

where z_{ij} represents the element at the ith row and jth column of $\boldsymbol{z}$. $\boldsymbol{\phi}(\boldsymbol{z})$ is the activation result of $\boldsymbol{z}$, and $\boldsymbol{z}$ is calculated by

$$\boldsymbol{z} = \boldsymbol{W} * \boldsymbol{\phi}(\boldsymbol{z}') + \boldsymbol{b}, \tag{2}$$

where $\boldsymbol{W}$ is the weight matrix, $\boldsymbol{b}$ is the bias term, $\boldsymbol{z}'$ is the outcome of the former convolution result, and $*$ is the symbol of convolution.

A similar procedure occurs in the last part of a network. Because there are six classes of land-cover types in this scene, we use a multi-class classification function to activate the outcome of the FC layers. The Softmax function (Bishop, 2006) (namely, a generalized-logistic function) computes the possibility, $\boldsymbol{p}(\boldsymbol{z})$, of each class as follows:

$$\boldsymbol{p}(\boldsymbol{z})_i = \frac{\mathrm{e}^{\boldsymbol{z}_i}}{\sum_{k=1}^{6} \mathrm{e}^{\boldsymbol{z}_k}}, \quad i = 1, 2, \ldots, 6, \tag{3}$$

where $\boldsymbol{z}$ is the output of FC layers.

In training, the outcomes of classification are compared with true labels. Errors are computed and propagated in reverse to adjust weights in the network (Rumelhart et al., 1988). This improves subsequent classification results. If we set $\boldsymbol{\omega}_{j+1}$ to be the updated weight matrix of $\boldsymbol{\omega}_j$, the weights are correlated as follows:

$$\boldsymbol{\omega}_{j+1} = \boldsymbol{\omega}_j - \mathrm{lr} \cdot \frac{\partial E}{\partial \boldsymbol{\omega}_j}, \tag{4}$$

where lr represents the learning rate; E is the total error calculated by

$$E = \frac{1}{2} \sum_{k=1}^{6} (t_k - o_k)^2, \tag{5}$$

where t_k is the true value, and o_k is the output value of the kth class.

2.4 Evaluation

Generally, the capability of a network is measured by training and test accuracy during training. However, in our study, the number of labeled pixels is far fewer than the complete set of pixels, causing difficulties in evaluating the output of the whole scene.

In that case, we use another criterion to estimate the predicted outcome (PA). Taking u_a as the number of unrecognized pixels of land-cover type a in area D, the unrecognized rate δ_a is calculated by

$$\delta_a = \frac{u_a}{n_D}, \tag{6}$$

where n_D is the number of pixels in area D.

2.5 Spectral weight

The unrecognized rate, δ_a, calculated in Section 2.4 describes the deficiency of the network, which means the network classifies these a-type pixels as other wrong types. With respect to the classifier, this means the possibilities of type a, calculated in Section 2.3, are fewer than the wrong types.

By simply adding an extra weight to the unrecognized type, δ of that type is reduced. However, because of the inequality of the weights that harm the stability and robustness of the network, δ of the other types increases.

By dividing the spectrum into visible light and infrared, we add unequal weights to these two parts.

Maintaining the sum of the weights of each type to be equal, the more sensitive part obtains a greater weight. In this circumstance, not only does the unrecognized rate decrease, but also the stability of the network is protected.

Additionally, the sensitivity is inversely proportional to δ. The lower δ is, the closer the PA is to the true label. Taking the wrongly recognized rate of type a in an area as γ_a, the spectral weight of type a is defined as

$$w_a = \left(\frac{0.243}{\delta_a + \varepsilon} + 0.374\right) \cdot \ln \frac{0.525}{\gamma_a + \varepsilon}, \tag{7}$$

where ε is a small number (with the denominator not zero), which is set as 1×10^{-5}; $\ln(0.525/(\gamma_a + \varepsilon))$ is a penalty term. A low unrecognized rate cannot be at the expense of a highly wrongly recognized rate. A penalty term maintains the balance between these two factors. Normalization is conducted as

$$\begin{aligned} W^{\mathrm{V}} &= \frac{w^{\mathrm{V}}}{w^{\mathrm{V}} + w^{\mathrm{IR}}}, \\ W^{\mathrm{IR}} &= \frac{w^{\mathrm{IR}}}{w^{\mathrm{V}} + w^{\mathrm{IR}}}, \end{aligned} \tag{8}$$

where W is the normalized spectral weight. Superscripts V and IR represent visible and infrared, respectively. The new possibility P is calculated by

$$P = W^{\mathrm{V}} p^{\mathrm{V}} + W^{\mathrm{IR}} p^{\mathrm{IR}}, \tag{9}$$

where p^{V} and p^{IR} are the two original probabilities.

3 Results and discussion

3.1 Spectral comparison

The spectral weight of each land-cover type defined by Eq. (8) is deduced from the comparisons among the outputs of the visible, infrared, and whole spectrum. During training, training and test accuracies of different inputs are compared (Fig. 3).

The curves in Fig. 3 show that, whether in training or testing, input from the whole spectrum has higher accuracy than the other two inputs, simply because the whole spectrum has more information than its two subsets.

Beyond this point, visible light input noticeably has a higher accuracy than infrared input. The inference is that more classes in those six land-cover types are more sensitive to visible light. While the proportion of the training set was reduced to 5%, test accuracies of visible light, infrared, and whole spectrum inputs decreased to 96.92%, 93.82%, and 97.44%, respectively.

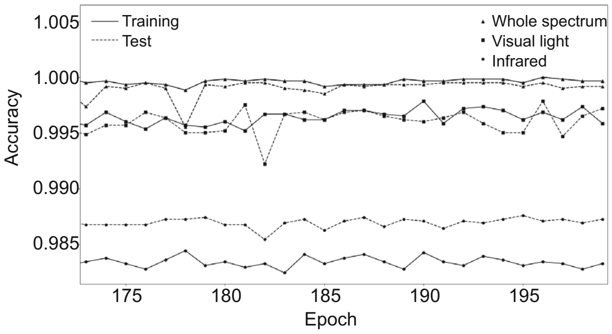

Fig. 3 Training and test accuracies of different inputs. The ratio of training/test samples is 1:1, which is the same as Tables 2 and 3

3.2 Manual correction

To calculate the spectral weight of each class, we selected, from every class, an area where the discrepancy between the PAs of the visible light and infrared outputs was the greatest (Fig. 4).

For buildings, roadways, concrete ground, arable land, vegetation, and water in Figs. 4a-4f, separately, we calculated the PAs and δ (Table 2).

Table 2 PAs and δ of different classes in different areas

(Area) Type	PA (%)		δ (%)	
	Visible	Infrared	Visible	Infrared
(a) Building	**91.38**	46.65	**0.60**	39.49
(b) Roadway	**97.32**	22.91	**2.68**	76.38
(c) C. ground	**87.25**	4.00	**12.71**	96.00
(d) A. land	86.56	**99.12**	0.11	0.14
(e) Vegetation	**90.35**	33.25	**3.18**	61.84
(f) Water	47.45	**99.41**	52.55	**0.04**

C. and A. are abbreviations for 'concrete' and 'arable'

It is clear that buildings, roadways, concrete ground, and vegetation are more sensitive in the visible band; whereas, arable land and water are more sensitive in the infrared band. This result corroborates the deduction from the learning curves in Section 3.1. Table 3 shows the W and w derived from Table 2.

3.3 Structure of improved network

According to Eq. (9), we designed an improved network, including two concurrent CNNs. Because

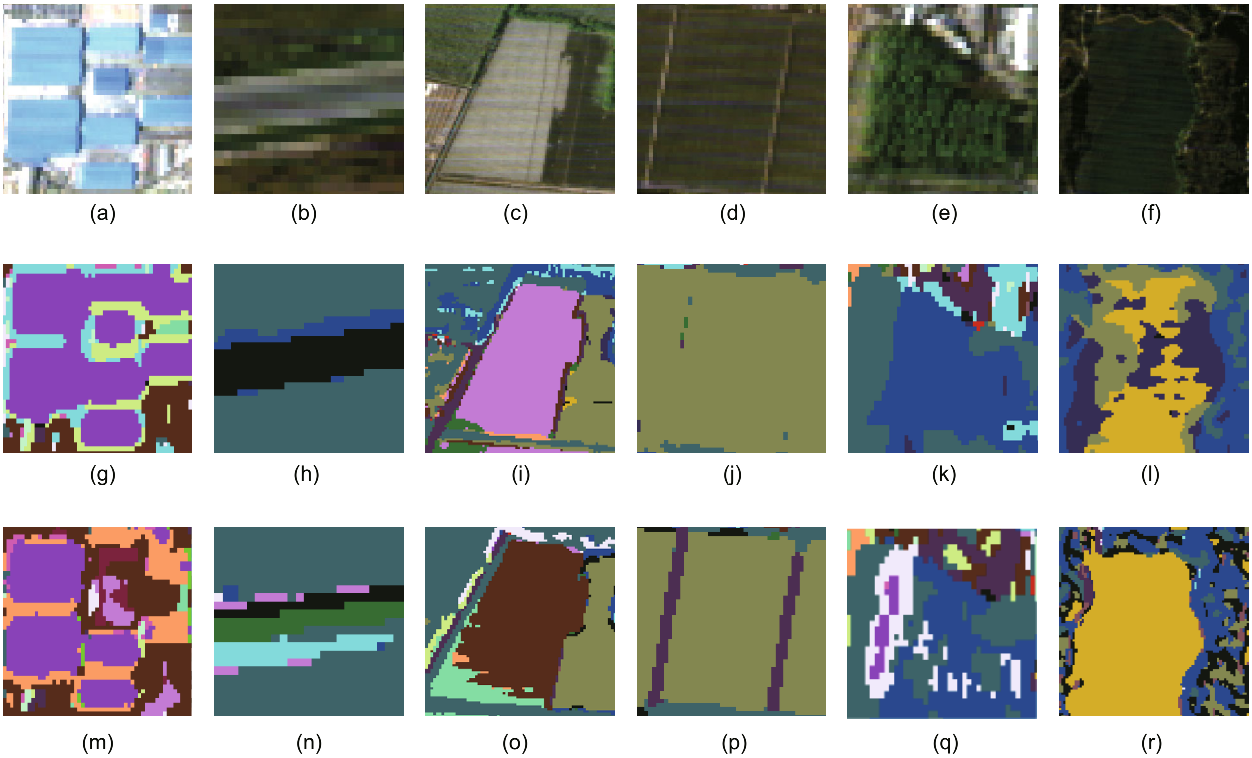

Fig. 4 Partial areas and the corresponding outputs of visible light and infrared inputs: (a)-(f) six areas in the original scene; (g)-(l) the same regions as (a)-(f) with visible light inputs; (m)-(r) the same regions as (a)-(f) with infrared inputs

Table 3 Spectral weights of different land-cover types in visible and infrared bands *

Weight	Building		Roadway		C. ground		A. land		Vegetation		Water	
	V.	I.	V.	I.	V.	I.	V.	I.	V.	I.	V.	I.
W	147.95	2.80	120.54	1.19	28.74	0.14	173.09	351.32	59.47	1.67	3.99	396.50
w	0.981	0.019	0.990	0.010	0.995	0.005	0.330	0.670	0.973	0.027	0.010	0.990

* Visible and infrared bands are abbreviated as V. and I.

there are 27 bands in the visible light region and 37 bands in the infrared, the numbers of convolution kernels in the first convolution layer are 27 in the visible part and 37 in the infrared part. In the spatial dimension, the scale of the kernels ranges from 2 × 2 to 4 × 4. Two layers of 3 × 3 kernels perform as well as one layer of 5 × 5 kernels (Szegedy et al., 2016), indicating that all the features in our input pictures can be extracted. The number of kernels decides the number of features to be extracted.

Finally, we created two convolution layers with 64 sets of 3 × 3 kernels in each layer and kept the number of channels consistent with the former outcome. In training, two CNNs are tuned; whereas, in predicting, spectral weight, w, is added to the probability calculations after activation. A sketch of the structure is shown in Fig. 5.

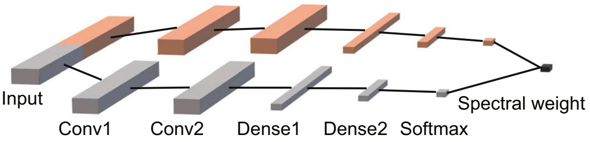

Fig. 5 Sketch of the improved network with spectral weight (Conv1, Conv2, Dense1, and Dense2 are the first convolutional layer, the second convolutional layer, the first dense layer, and the second dense layer, respectively)

3.4 Comparison with whole spectrum input

Results show that the advantage of the two original networks, synthesized in the improved network,

Table 4 Accuracies in public data sets

Type	Accuracy in KSC data set (%)			Accuracy in Botswana data set (%)		
	Whole spectral	Proposed	Improvement	Whole spectral	Proposed	Improvement
1	96.45	99.60	3.15	100.00	100.00	0.00
2	95.47	98.76	3.29	100.00	100.00	0.00
3	94.53	85.93	-8.60	100.00	100.00	0.00
4	93.25	88.89	-4.36	100.00	100.00	0.00
5	91.30	80.74	10.56	94.79	98.88	4.09
6	82.09	96.50	14.41	98.88	99.25	0.37
7	99.04	99.04	0.00	100.00	100.00	0.00
8	97.91	98.14	0.23	100.00	100.00	0.00
9	99.42	99.80	0.38	94.90	98.40	-3.50
10	100.00	100.00	0.00	98.79	99.59	0.80
11	100.00	100.00	0.00	98.68	100.00	1.32
12	99.76	100.00	0.24	100.00	100.00	0.00
13	100.00	100.00	0.00	100.00	100.00	0.00
14				100.00	100.00	0.00
Total	95.30	96.84	1.54	99.24	99.45	0.21

20% of samples were chosen as training sets, while 30% as validation sets, 50% as test sets

is such that they became more sensitive to all types in a scene. To quantify the capability of the improved network, we compared the new output with the output of the input from the whole spectrum. A representative area was chosen to calculate the PA. In this area, the PA of the improved network is 98.32%; whereas the PA of the whole spectrum input is 96.89%. Improvements in the 5%, 10%, 20%, 30%, and 40% training proportions are 1.44%, 1.39%, 1.51%, 1.47%, and 1.33%, respectively. These results prove that spectral weight promotes sensitivity to some objects on the earth's surface, and the improvements are almost independent of the proportion of the training sets.

3.5 Results of public data sets

In whole images in the KSC data set, we achieved overall accuracy (OA) of 95.30%, 94.90%, and 86.19% with whole spectrum, visible light, and infrared, respectively. With the proposed method, the OA of the KSC data set increased to 96.84%, 1.54% higher than the whole spectrum input. Almost every type was classified more accurately after applying the proposed method (Table 4). In the Botswana data set, the OA of whole spectrum, visible light, infrared, and the proposed method are 99.24%, 96.99%, 99.21%, and 99.45%, respectively. According to Table 4, every type in the proposed method is greater or equal to the whole spectrum input.

4 Conclusions

We proposed a CNN structure constructed by two concurrent CNNs. Taking advantage of the high spectral resolution of hyperspectral images, we found the spectral dependencies of land-cover types in a scene by manually calculating the unrecognized and wrongly recognized rates. This correlation enabled us to develop a spectral weight matrix, which adds spectral dependencies, as a key factor, into the classification stage. Compared with the well-tuned whole spectrum input structure, both results in our data and public data sets show that the spectral weight, based on spectral sensitivity, helps increase the classification accuracy and enhances the sensitivity of the network to all types in this scene.

Contributors

Cheng-ming YE designed the research. Xin LIU and Hong XU processed the corresponding data. Xin LIU and Cheng-ming YE wrote the first draft of the manuscript. Jonathan LI and Hong XU helped organize the manuscript. Xin LIU, Yao LI, Shi-cong REN, and Jonathan LI revised and edited the final version.

Conflict of interest

Cheng-ming YE, Xin LIU, Hong XU, Shi-cong REN, Yao LI, and Jonathan LI declare that they have no conflict of interest.

References

Bengio Y, Courville A, Vincent P, 2013. Representation learning: a review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798-1828.
https://doi.org/10.1109/TPAMI.2013.50

Bioucas-Dias JM, Plaza A, Camps-Valls G, et al., 2013. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine*, 1(2):6-36.
https://doi.org/10.1109/MGRS.2013.2244672

Bishop CM, 2006. Pattern Recognition and Machine Learning. Springer, New York, USA.

Cai Y, Yang ML, Li J, 2015. Multiclass classification based on a deep convolutional network for head pose estimation. *Frontiers of Information Technology & Electronic Engineering*, 16(11):930-939.
https://doi.org/10.1631/FITEE.1500125

Cao JY, Chen Z, Wang B, 2016. Deep convolutional networks with superpixel segmentation for hyperspectral image classification. Proceedings of 2016 IEEE International Geoscience and Remote Sensing Symposium, p.3310-3313.
https://doi.org/10.1109/IGARSS.2016.7729856

Chen YS, Zhao X, Jia XP, 2015. Spectral-spatial classification of hyperspectral data based on deep belief network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2381-2392.
https://doi.org/10.1109/JSTARS.2015.2388577

Du B, Zhang LP, 2014. A discriminative metric learning based anomaly detection method. *IEEE Transactions on Geoscience and Remote Sensing*, 52(11):6844-6857.
https://doi.org/10.1109/TGRS.2014.2303895

Glorot X, Bordes A, Bengio Y, 2011. Deep sparse rectifier neural networks. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, p.315-323.

Gu YF, Chanussot J, Jia XP, et al., 2017. Multiple kernel learning for hyperspectral image classification: a review. *IEEE Transactions on Geoscience and Remote Sensing*, 55(11):6547-6565.
https://doi.org/10.1109/TGRS.2017.2729882

Hinton GE, Salakhutdinov RR, 2006. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504-507.
https://doi.org/10.1126/science.1127647

Hotelling H, 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417-441.
https://doi.org/10.1037/h0071325

Krizhevsky A, Sutskever I, Hinton GE, 2017. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84-90.
https://doi.org/10.1145/3065386

LeCun Y, Boser B, Denker JS, et al., 1989. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541-551.
https://doi.org/10.1162/neco.1989.1.4.541

Li J, Bioucas-Dias JM, Plaza A, 2010. Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4085-4098.
https://doi.org/10.1109/TGRS.2010.2060550

Li YS, Xie WY, Li HQ, 2017. Hyperspectral image reconstruction by deep convolutional neural network for classification. *Pattern Recognition*, 63:371-383.
https://doi.org/10.1016/j.patcog.2016.10.019

McCulloch WS, Pitts W, 1943. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115-133.
https://doi.org/10.1007/BF02478259

Melgani F, Bruzzone L, 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778-1790.
https://doi.org/10.1109/TGRS.2004.831865

Mou LC, Ghamisi P, Zhu XX, 2017. Deep recurrent neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7):3639-3655.
https://doi.org/10.1109/TGRS.2016.2636241

Onojeghuo AO, Blackburn GA, Huang JF, et al., 2018. Applications of satellite 'hyper-sensing' in Chinese agriculture: challenges and opportunities. *International Journal of Applied Earth Observation and Geoinformation*, 64:62-86.
https://doi.org/10.1016/j.jag.2017.09.005

Rosenblatt F, 1958. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386-408.
https://doi.org/10.1037/h0042519

Rumelhart DE, Hinton GE, Williams RJ, 1988. Learning internal representations by error propagation. *In*: Collins A, Smith EE (Eds.), Readings in Cognitive Science. Morgan Kaufmann, San Mateo, USA.
https://doi.org/10.1016/B978-1-4832-1446-7.50035-2

Sigurdsson J, Ulfarsson MO, Sveinsson JR, 2014. Semi-supervised hyperspectral unmixing. Proceedings of 2014 IEEE Geoscience and Remote Sensing Symposium, p.3458-3461.
https://doi.org/10.1109/IGARSS.2014.6947226

Szegedy C, Vanhoucke V, Ioffe S, et al., 2016. Rethinking the inception architecture for computer vision. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, p.2818-2826.
https://doi.org/10.1109/CVPR.2016.308

Wang LD, Zhou W, Xing Y, et al., 2019. A novel method based on convolutional neural networks for deriving standard 12-lead ECG from serial 3-lead ECG. *Frontiers of Information Technology & Electronic Engineering*, 20(3):405-413.
https://doi.org/10.1631/FITEE.1700413

Wentz EA, Anderson S, Fragkias M, et al., 2014. Supporting global environmental change research: a review of trends and knowledge gaps in urban remote sensing. *Remote Sensing*, 6(5):3879-3905.
https://doi.org/10.3390/rs6053879

Yang XH, Huang JF, Wang FM, et al., 2006. A modified chlorophyll absorption continuum index for chlorophyll estimation. *Journal of Zhejiang University-SCIENCE A*, 7(12):2002-2006.
https://doi.org/10.1631/jzus.2006.A2002

Ye CM, Cui P, Pirasteh S, et al., 2017. Experimental approach for identifying building surface materials based on hyperspectral remote sensing imagery. *Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering)*, 18(12):984-990. https://doi.org/10.1631/jzus.A1700149

Zhang F, Du B, Zhang LP, et al., 2016. Hierarchical feature learning with dropout k-means for hyperspectral image classification. *Neurocomputing*, 187:75-82. https://doi.org/10.1016/j.neucom.2015.07.132

Zhao WZ, Du SH, 2016. Learning multiscale and deep representations for classifying remotely sensed imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 113:155-165. https://doi.org/10.1016/j.isprsjprs.2016.01.004

Zhong P, Gong ZQ, Li ST, et al., 2017. Learning to diversify deep belief networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(6):3516-3530. https://doi.org/10.1109/TGRS.2017.2675902

Zhong ZL, Li J, Luo ZM, et al., 2018. Spectral-spatial residual network for hyperspectral image classification: a 3-D deep learning framework. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2):847-858. https://doi.org/10.1109/TGRS.2017.2755542

中文概要

题　目：基于卷积神经网络和光谱敏感度的高光谱影像分类方法

目　的：由于高光谱成像的特性，高光谱遥感影像较光学、多光谱影像具有更多的光谱信息，因此对高光谱影像地物的分类也相对困难。为提高分类精度，本文提出一个新的高光谱遥感影像分类模型。

创新点：考虑到不同的地物覆盖对不同波段范围的电磁波有不同的敏感度，本文提出一个基于卷积神经网络和光谱敏感度的深度学习模型，以提高对高光谱遥感影像地物分类的准确率。通过在最终的分类器后添加一个光谱权重，该模型能够更准确地分类地物。

方　法：1. 将带标记的样本在光谱维度上分为可见光和红外波段，并将部分样本作为训练集和测试集输入到网络中进行训练。2. 训练完成后利用模型对全图进行预测，并通过部分预测结果计算出未识别率δ和误识别率γ两个参数。3. 利用δ和γ可计算出不同光谱范围的光谱权重并将其置于分类器前（图5）。

结　论：1. 模型加入光谱权重后的分类准确率较之前提高了约2%。2. 利用公共数据集测试后显示，使用了光谱权重的卷积神经网络模型的分类精度比未使用光谱权重的模型高约1%。3. 本文结果显示，利用不同地物对电磁波的敏感性差别可以增加不同地物间的差异，从而提升分类模型的性能。

关键词：高光谱影像；深度学习；卷积神经网络；光谱灵敏度