

Deep learning classification of coastal wetland hyperspectral image combined spectra and texture features: A case study of Huanghe (Yellow) River Estuary wetland

Yabin Hu^{1, 2}, Jie Zhang², Yi Ma^{2*}, Xiaomin Li², Qinpei Sun³, Jubai An¹

¹ Information Science and Technology College, Dalian Maritime University, Dalian 116026, China

² First Institute of Oceanography, Ministry of Natural Resources, Qingdao 266061, China

³ College of Surveying Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China

Received 14 March 2018; accepted 2 May 2018

© Chinese Society for Oceanography and Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

This paper develops a deep learning classification method with fully-connected 8-layers characteristics to classification of coastal wetland based on CHRIS hyperspectral image. The method combined spectral feature and multi-spatial texture feature information has been applied in the Huanghe (Yellow) River Estuary coastal wetland. The results show that: (1) Based on testing samples, the DCNN model combined spectral feature and texture feature after K-L transformation appear high classification accuracy, which is up to 99%. (2) The accuracy by using spectral feature with all the texture feature is lower than that using spectral only and combining spectral and texture feature after K-L transformation. The DCNN classification accuracy using spectral feature and texture feature after K-L transformation was up to 99.38%, and the outperformed that of all the texture feature by 4.15%. (3) The classification accuracy of the DCNN method achieves better performance than other methods based on the whole validation image, with an overall accuracy of 84.64% and the Kappa coefficient of 0.80. (4) The developed DCNN model classification algorithm ensured the accuracy of all types is more balanced, and it also greatly improved the accuracy of tidal flat and farmland, while kept the classification accuracy of main types almost invariant compared to the shallow algorithms. The classification accuracy of tidal flat and farmland is up to 79.26% and 56.72% respectively based on the DCNN model. And it improves by about 2.51% and 10.6% compared with that of the other shallow classification methods.

Key words: coastal wetland, hyperspectral image, deep learning, classification

Citation: Hu Yabin, Zhang Jie, Ma Yi, Li Xiaomin, Sun Qinpei, An Jubai. 2019. Deep learning classification of coastal wetland hyperspectral image combined spectra and texture features: A case study of Huanghe (Yellow) River Estuary wetland. *Acta Oceanologica Sinica*, 38(5): 142–150, doi: 10.1007/s13131-019-1445-z

1 Introduction

Coastal wetland locates in a transitional belt between land and marine system. It has the ability to reserve resources, regulate climate, protect environment and purify air, while can provide good habitat for wild animals and plants. Coastal wetland is connected with human life closely. In recent years, its area has declined dramatically due to human activities and natural factors, which disturbed the balance of land-sea resources exchange and affected human being directly. Therefore, it is significant to carry out distribution and change monitoring of the coastal wetland, which is beneficial to the effective utilization and sustainable development of wetland resources and environment.

Remote sensing technology has the advantage of large area synchronous observation and high time resolution, which solves the problem that people are hard to enter the coastal wetland. It is a challenge to monitor the distribution and variation of the wetland types accurately. In recent years, many scholars have employed different methods to carry out remote sensing classifica-

tion research based on multi-spectral and multi-temporal satellite remote sensing images, including support vector machine (SVM) (Melgani and Bruzzone, 2004), decision tree (DT) (Licciardi et al., 2009), decision fusion (Li et al., 2015; He et al., 2010), random forest (RF) and fuzzy mathematics etc (Xu et al., 2006; Freund, 1995; Chubey et al., 2006; Sun et al., 2013; Cao et al., 2016). However, it is difficult to obtain satisfied results in real scene. In fact, the above-mentioned coastal wetland remote sensing classification methods are mainly applied with multi-spectral remote sensing images of medium and low resolution. These application methods are confined to the level of shallow learning, and it is difficult to meet the requirement of fine classification for coastal wetland. The deep learning is a new research direction in the field of machine learning, whose basic process is to obtain a deep network structure with multiple layers through a certain training model (Hinton et al., 2014), and it solves the strong nonlinear problem that shallow methods are out of operation. At present, deep learning has obtained great achievements in the fields of image processing (Krizhevsky et al., 2012; Farabet

Foundation item: The National Natural Science Foundation of China under contract No. 61601133 and 41206172; the Marine Application System of High Resolution Earth Observation System Major Project.

*Corresponding author, E-mail: mayimail@fio.org.cn

et al., 2013), natural language recognition (Waibel et al., 1989; Hinton et al., 2012) and target detection (Xu et al., 2016; Teoh et al., 2012; Tian et al., 2016). It has become a hotspot in the hyperspectral remote sensing images classification (Li et al., 2012; Mei et al., 2016; Yang et al., 2016; Chen et al., 2017; Lee and Kwon, 2016; Slavkovikj et al., 2015; Tarabalka et al., 2009). Convolution Neural Network (CNN) (Hinton and Salakhutdinov, 2006), a deep learning method, has the characteristics of image displacement and rotation invariance. Yue et al. (2015) proposed a CNNs-LR model based on spectral-spatial feature, then applied the model in the Pavia data set. The results showed that the overall accuracy of classification was 95.18%, and the Kappa coefficient was 93.64% (Yue et al., 2015). Hu et al. (2015) used 5-layer CNNs model with hyperspectral image spectral features to carry out classification, the results indicated that the method was superior to the traditional SVM method (Hu et al., 2015). In summary, many scholars used CNNs algorithm in hyperspectral image remote sensing classification, but the research stayed in the verification stage of model algorithm generally. The accuracies of these methods were based on testing samples only, lacking global region verification. On the other hand, the DCNN model has not been applied in the coastal wetland remote sensing image classification. In fact, it is necessary to obtain the high accuracy coastal wetland classification results by using the DCNN model.

According to the needs of coastal wetland monitoring, the paper put forward a deep learning method with fully-connected 8-layers characteristics to carry out coastal wetland classification based on CHRIS hyperspectral remote sensing image, and we applied it in Huanghe (Yellow) River Estuary coastal wetland. Simultaneously, we compared the remote sensing classification accuracy between the DCNN method and several shallow learning methods.

2 Data and methods

2.1 Data

2.1.1 Hyperspectral remote sensing data

The hyperspectral image was used in this paper acquired from PROBA satellites launched by ESA on October 22, 2001. The satellite was equipped with CHRIS (Compact High Resolution Imaging Spectrometer), and the revisit period was 18 d. CHRIS Hyperspectral images with 18 bands can be obtained from five imaging modes and five angle observations. There were 18 bands in each image, whose spectral range covered from visible spectrum to near infrared spectrum. Their wavelength of spectral bands was ranging from 406 to 1 036 nm. The width of each image width was 13 km×13 km, and the spectral resolution was ranging from 5.8 to 44.1 nm, while spatial resolution was 17 m. The detail information of band parameter setting (An et al., 2016) was shown in Table 1. The CHRIS hyperspectral images of coastal wetland was used in this paper that were obtained in June 2012 and September 2009, whose operating mode stays at State 2 and image angle is 0°. The preprocessing of CHRIS coastal wetland image includes geometric correction, missing pixel filling and noise elimination.

In this paper, the Huanghe River Estuary coastal wetland, the study area, is a national nature reserve surrounded by the Bohai Bay and Laizhou Bay, locates in Dongying City, Shandong Province. It is the habitat for many rare and endangered species. The geographical coordinates of the study area are between 37°44'–37°51'N and 119°04'–119°14'E, which belongs to the warm temperate continental monsoon climate, and the annual average temperature and precipitation is about 12.8°C and 555 mm, respectively. The location of the study area is shown in Fig. 1.

2.1.2 Field data

The study area was covering the intersection of the old and

Table 1. The band parameter setting of CHRIS hyperspectral image

Band parameter	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7	Band 8	Band 9
Band center/nm	411.3	443.6	491.8	511.5	532.0	563.6	576.1	593.2	625.3
Band width/nm	10.9	10.6	11.7	12.9	11.6	13.8	11.0	16.0	13.7
Band parameter	Band 10	Band 11	Band 12	Band 13	Band 14	Band 15	Band 16	Band 17	Band 18
Band center/nm	654.3	672.9	684.1	692.7	710.7	760.4	786.1	878.6	1 026.7
Band width/nm	15.4	10.9	11.4	5.8	18.5	14.2	22.8	27.6	44.1

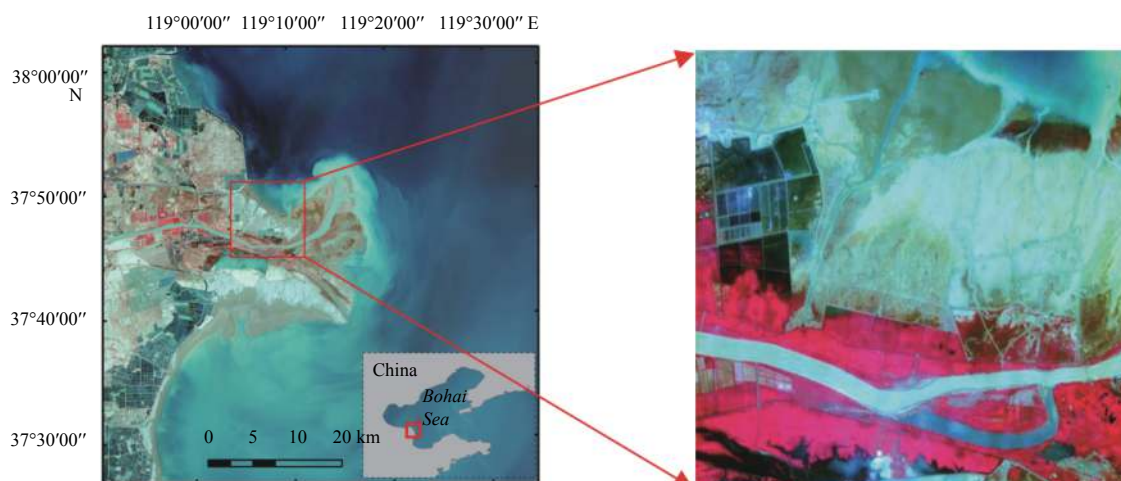


Fig. 1. The location of the Huanghe River Estuary coastal wetland.

new rivers in the Huanghe River Estuary, which was situated at the intersection of the two branches of Qingbacha River and Qingshuigou River. The study area included natural wetlands (reed, Tamarix, spruce and seawater) and artificial wetlands (farmland, aquaculture and ponds). In order to evaluate the classification accuracy of CHRIS image accurately, our group went to the study area to carry out field survey by using a typical sample and route record method in June 2008 and September 2012.

We obtained hundreds of hyperspectral data and more than 400 photos of typical types. Based on field data and high-resolution remote sensing images in the same period, we acquired the interpretation map of typical types by using human-computer interaction method, and we took it as a whole validation image (Fig. 2). At the same time, we verified the accuracy of the interpretation map through RTK instruments in May 2016 (Fig. 3).

Convolution Neural Network (CNN) is a kind of deep learning method, which is specially designed for image classification and recognition based on multi-layer neural network. A typical CNN model includes input layer, convolution layer, drop sampling layer, fully-connected layer and output layer.

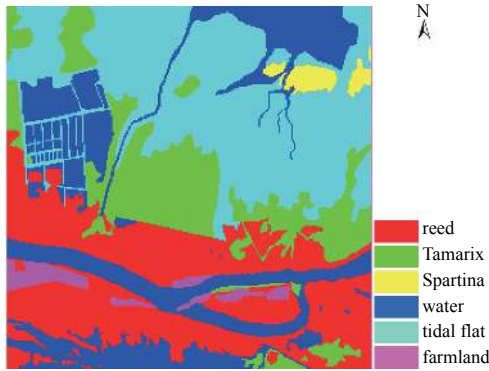


Fig. 2. The interpretation map of the Huanghe River Estuary coastal wetland.



Fig. 3. The photo of field survey.

2.1.3 DCNN structure

Although the hyperspectral remote sensing is characterized by high spectral resolution, the hyperspectral image sometimes indicates the phenomenon that different types with same spectrum and the same types with different spectrum due to the complicated atmospheric, light and sensor conditions. It will affect the image classification accuracy seriously if we just adopt the spectral characteristics. Therefore, we introduced the spatial fea-

ture information of the similarity of adjacent pixels into the classification algorithm. The developed classification algorithm was combined with spatial features and spectral features, which took the direction and neighborhood information into account, and it might effectively improve the classification accuracy of object types. Although many scholars used spectral features and spatial features to identify types based on hyperspectral images, the coastal wetland DCNN model classification based on hyperspectral remote sensing image is relatively poor. Therefore, this paper proposed an 8-layers DCNN model that took spatial features and spectral features into account, which consisted of an input layer, two convolution layers, two drop sampling layers, two fully-connected layers and one output layer. The proposed DCNN model structure was shown in Fig. 4.

2.1.4 Input layer

The input layer of the proposed DCNN model consists of a feature vector which contains the spectral and texture features information. The input vector z_i can be expressed by Eq. (1):

$$z_i = [X_i^1, X_i^2, \dots, X_i^d, X_i^1, y_i^1, y_i^2, \dots, y_i^d] \quad i \in (1, 2, \dots, N), \quad (1)$$

where z_i denotes the feature value of image pixels, N is the number of samples, d is the number of bands, and X_i^1 indicates the spectral value of the i sample in Band d .

The input vector g^{l+1} of convolution layer is the result of the conversion of input layer by Eq. (2):

$$g^{l+1} = f^l(W^{lT}g^l + b^l), \quad (2)$$

where g^l denotes the input value of l th layer and the output value of $(l-1)$ th layer, W^T is a weight matrix of the l th layer acting on the input feature data, b^l is an additive bias vector for the l th layer, and $f^l(\bullet)$ represents the activate function of the l th layer.

2.1.5 Convolution layer

Convolution layer is a process to obtain a new feature map that a result of the input feature map can obtain through convolution operation with the learning of the convolution kernel and an activation function. A DCNN model can include multiple convolution layers. According to different purposes, different convolution can be used to convolve the input feature map in the convolution layer, so different convolution kernels will obtain different feature map by convolution operation. An output feature map can be obtained by multiple convolution processes. Multi-convolution kernel is a sampling process in time domain or space domain, which can reduce the complexity of the network and the resolution of the feature map effectively, and reduce the sensitivity to displacement, rotation and scaling. The convolution layer is calculated by Eq. (3):

$$V_m^l = f_l \left(\sum_{n \in G} V_n^{l-1} \otimes H_{mn}^l + b_m^l \right), \quad (3)$$

where V_m^l is the activate value of the output feature map in the l th layer, G represents the feature map in l th feature layer, H_{mn}^l represents the convolution kernel that links the input feature map n in l th feature layer and the output feature map m in $(l-1)$ th feature layer, b_m^l is the bias coefficient for the l th feature layer, \otimes denotes convolution operation, $f_l(\bullet)$ represents the

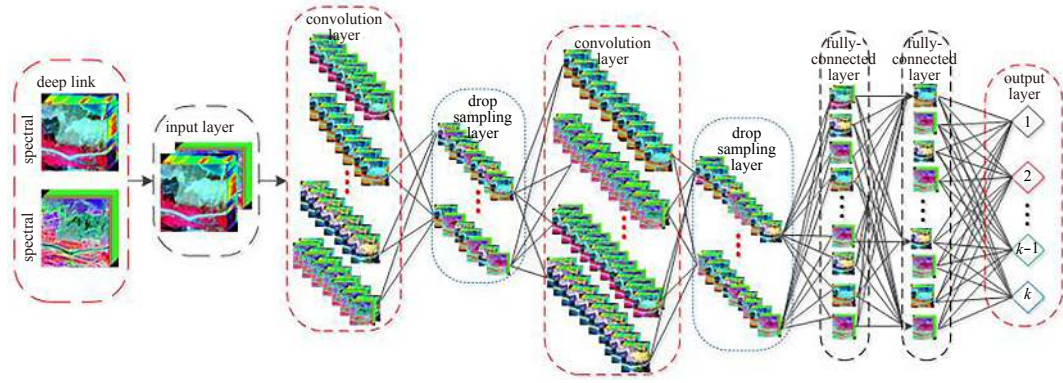


Fig. 4. The structure of the proposed DCNN model.

softmax activate function. The softmax regression models is defined as Eq. (4):

$$f(x) = \frac{1}{1 + e^{-ax}}, \quad (4)$$

where a is the tilt coefficient.

2.1.6 Drop sampling layer

The main purpose of drop sampling layer is to reduce the number of output parameters of feature map through the convolution layer, and to feature that can make the feature representation keep the invariance of rotation, translation and stretching. If the number of feature maps is too large in the DCNN model, it will lead to overfitting. The sampling layer is calculated by Eq. (5):

$$V_m^l = f^l(\alpha_m^l \times \text{down}(V_m^{l-1}) + b_m^l), \quad (5)$$

where α_m^l and b_m^l represent the multiplicative bias coefficient and the additive bias coefficient of feature map m for the l th feature layer, respectively; and $\text{down}(\bullet)$ is a down-sampling function acting on the l th feature layer.

Drop sampling methods include max-pooling and mean pooling. The max-pooling method is adopted in the proposed DCNN model, which is to reduce the estimated mean shift caused by convolutional parameters error to preserve image texture information, and the result of error is obtained by the statistical calculation of the features of different positions. The down-sampling value is given by Eq. (6):

$$U_m = \max(U_m^1, U_m^2, \dots, U_m^q), \quad (6)$$

where U_m^q represents the value of element q in feature map m ; and q is the number of feature map m , which depends on the size of the down-sampling window.

2.1.7 Fully-connected layer

The fully-connected layer is a reorganization process of the feature map in the model, which can improve the effect of model training. Compared with the sparse connection, the full connection takes the nature of all feature maps in upper layer network into account, and obtains the combined coefficient of the feature maps by learning method, which further improves the expression ability and adaptability of model. The value of the feature map V_n^l through fully-connected layer can be expressed by the

Eq. (7):

$$V_n^l = f\left(\sum_{m=1}^Z \partial_{mn} * V_m^{l-1} + b_n^l\right), \quad (7)$$

where Z is the total number of feature maps in the input $(l - 1)$ th feature layer, ∂_{mn} represents the weight value that link the feature map n in l th feature layer and the input feature map m in $(l - 1)$ th feature layer, b_n^l represents the additive bias coefficient of feature map n for the l th feature layer, the weight ∂_{mn} can be enhanced by the soft-max function of a set of unconstrained implicit weight c_{mn} , and ∂_{mn} is defined as Eq. (8):

$$\partial_{mn} \Theta \begin{cases} 0 \leq \partial_{mn} \leq 1, \\ \partial_{mn} = \frac{\exp(c_{mn})}{\sum_k \exp(c_{kn})}, \\ \sum_m \partial_{mn} = 1, \end{cases} \quad (8)$$

where Θ denotes constraint condition.

2.1.8 Output layer

The output layer of the proposed DCNN model needs to predict the maximum likelihood of input neurons, and a neuron in output layer reflects a class of object. The softmax regression activation function was used to enhance the binding between neurons in the output layer, and to reduce the confusion of the categories of neurons. The class estimation function can be defined by the softmax regression activation function P as Eq. (9):

$$P = \begin{bmatrix} P(Y = 1 | V, W, b) \\ P(Y = 2 | V, W, b) \\ \dots \\ P(Y = k | V, W, b) \end{bmatrix} = \frac{1}{\sum_{i=1}^k e^{W_{nk}^T V_n + b_k}} \begin{bmatrix} W_{n1}^T V_n + b_1 \\ W_{n2}^T V_n + b_2 \\ \dots \\ W_{nk}^T V_n + b_k \end{bmatrix}, \quad (9)$$

where W and b represents the multiplicative weight and the additive bias coefficient; W_{nk} and b_k represent the weight and bias

coefficient of the input neuron n which is classified as class k of object type; Y represents the classification result of the input neuron, which corresponds to the label of training samples; and $P(\bullet)$ represents the potential probability of input pixel g is attributable to object type k . The probability of sample classification i in the softmax function is given by Eq. (10):

$$P = (Y = i | V, W, b) = \frac{e^{W_{nk}^T V_n + b_k}}{\sum_{i=1}^k e^{W_{nk}^T V_n + b_k}}. \quad (10)$$

2.2 Feature extraction

2.2.1 Spectral feature

The spectral curve reflects the characteristics of absorption and reflection of objects. Different objects indicate different spectral waveform. Spectral features can be extracted from spectral curve, such as absorption wavelength position, reflection value, depth, width, slope, symmetry, area and so on. Based on the spectral feature of parameters, the purposes of rapid identification and accurate classification of different object types can be achieved.

The types of the Huanghe River Estuary coastal wetland are diverse. According to the features of CHRIS hyperspectral image and field survey data, the types of Huanghe River Estuary coastal wetland are defined as six categories: reed, Tamarix, Spartina, water, tidal flat and farmland.

2.2.2 Texture feature

Texture is an inherent important feature in an image, which is formed by the gray distribution of different objects in the spatial position, so there will be a certain degree of gray relationship among the adjacent pixels in image space. Of course, different object types indicate different texture features in the image. Texture Gray Level Co-occurrence Matrix (GLCM) is an algorithm that describes the properties of gray spatial correlation to reflect texture features, which reflects information of image gray level, neighborhood pixel spacing, direction and amplitude of change. The GLCM will be different due to the diversity of texture scale, so we use the texture gray level covariance matrix algorithm to extract eight texture features in the CHRIS hyperspectral images of coastal wetland. These texture features consist of mean, variance, homogeneity, contrast, dissimilarity, entropy, angular second moment and correlation.

The eight types of texture features based on the GLCM algorithm contain different direction information and neighborhood information, but not all the texture features are beneficial to the classification accuracy, some texture features may reduce the image classification accuracy. Therefore, it is necessary to carry out an information comprehensive selection from the 8 texture features, which can preserve the texture information and improve the efficiency and the classification accuracy at the same time. In this paper, K-L transform (Karhunen-Loeve transform) is used to extract the information of 8 types of texture features, which is a best transformation with strong correlation and mean square error based on the statistical properties. The first component after K-L transformation has the most of texture information, and the texture information of the other components decreases in turn. The amount of texture information contained in the first three components is up to 95%, so we retain them as the texture feature information.

2.3 Model training

2.3.1 CNN parameter setting

The input of the proposed DCNN model included spectral features and texture features. The fusion result contained 21-dimensional information, because we obtained the first three components of texture features with 8 texture parameters after K-L transform. In order to facilitate the operation of DCNN model, we expanded the 21-dimensional information twice, namely, the input layer contains 42-dimensional information. In this paper, the parameters of the DCNN model were set as follows: the number of feature maps was 4 and 8, respectively, the convolution kernel size was 3×3, and the number of iterations was 10.

2.3.2 The label of training samples and testing samples

In order to achieve the fine classification of object types of coastal wetland, we need to train and test the DCNN model to find the best matching parameters. The selection of training samples and test samples affected the accuracy of classification greatly. Because the characteristic of non-singularity about training sample pixels would change the construction of nonlinear functions in DCNN model, and then confuse the attribution of pixels. Therefore, the selection of typical training samples and test samples must ensure the purity of pixels. The selection of test samples was used to evaluate the accuracy of the DCNN model established by the training samples. In this paper, we selected 6 classes of typical object types as training samples and test samples based on field data and pictures in the Huanghe River Estuary, which were reed, Tamarix, Spartina, water, tidal flat and farmland. At last we selected 7 092 training samples and 1 396 test samples, respectively. The sample distribution was shown in Fig. 5 and Table 2.

We performed 10 Monte Carlo runs, where each run we selected a training set and a testing set of the labeled samples according to Table 2 to train our model. We reported the average and standard error of the 10 Monte Carlo runs in terms of the overall

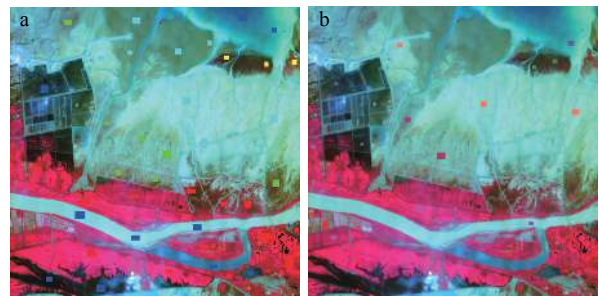


Fig. 5. The sample distribution map. a. Testing samples and b. training samples.

Table 2. The information of training samples and testing samples

Type	Training samples	Testing samples
Reed	2 428	251
Tamarix	1 000	322
Spartina	250	140
Water	1 828	242
Tidal flat	1 349	284
Farmland	237	157
Sum	7 092	1 396

classification accuracy (OCA).

3 Results and discussion

3.1 Classification accuracy evaluation of DCNN model based on testing samples

Based on the six typical types of training samples and testing samples selected from the coastal wetland in the Huanghe River Estuary, we carried out the classification experiments by the developed DCNN model in the case of using the spectral feature information only, combining the spectral feature information with all the texture feature information, and combining the spectral feature information with the texture feature after K-L transformation. At the same time, this paper introduced support vector machine (SVM) to compare with the proposed DCNN model. SVM is a machine learning method that can solve the existing problems such as small sample, nonlinearity, high dimension and local minimum point. The core of SVM is to find an optimal separable hyperplane, which distinguishes two types of samples from the training sample. This paper uses the SVM algorithm based on 4 types of kernel function called linear, polynomial, radial basis function and sigmoid kernel function. we carried out the classification experiments of CHRIS hyperspectral image in the case of using spectral feature information only and combining the spectral feature information with the texture feature after K-L transformation by the same training samples. Then we evaluated the classification accuracy by the same testing samples. The test results based on different classification methods are shown in Table 3.

It can be seen from Table 3, the classification accuracy of reed in all kinds of methods is up to 100%. The classification accuracy of Tamarix by SVM method based on linear and polynomial kernel was up to 100%. The classification accuracy of water by SVM method based on the polynomial and RBF kernel, and the DCNN model by using the spectral feature information combined with the K-L transformed texture feature information reaches 100%. While the classification accuracy of tidal flat reaches up to 99%, only in the case of combing spectral feature with all the texture feature classified by the DCNN model. In the form of different classification methods, the DCNN model combing spectral feature with texture K-L transformation feature had the highest classification accuracy of farmland, reaching 98.41%, and the accuracy of SVM algorithm based on sigmoid kernel is the lowest. Among all the classification methods, the overall accuracy of testing sample was up to 99% based on the DCNN model combined spectral feature and texture feature after K-L transformation only.

At the same time, it can be found in Table 3 that in a classification method, the accuracy by using spectral feature with all the texture feature was lower than that using spectral only and spec-

tral combing texture feature after K-L transformation, and the accuracy by using spectral feature only was lower than that combing spectral feature and texture feature after K-L transformation. In each classification method, compared with the two cases that based on spectral feature and combined spectral feature and all the texture feature, the classification accuracy of combined spectral feature and texture feature after K-L transformation improves by about 2% and 3%. In the three different cases by DCNN model, we could see that the classification accuracy by using all texture feature was the lowest, only 95.23%, and the classification accuracy by using spectral feature and texture feature after K-L transformation was up to 99.38%, and it improved by 4.15% compared with using all the texture feature. The method used all the texture features will affect the accuracy of image classification, and the method used texture feature after K-L transformation. It can also be explained that not all texture features are beneficial to the classification results, some of them will affect the accuracy of image classification.

3.2 Classification accuracy evaluation of DCNN model based on whole validation image

In this study, we selected five classification methods whose classification accuracy was over 98% by being applied to the CHRIS high resolution image about the Huanghe River Estuary. The total pixel number of CHRIS hyperspectral data is 262 144. The results of the five classification methods were shown in Fig. 6. In order to verify the accuracy of five classification methods for coastal wetland, we used the remote sensing interpretation of the Huanghe River Estuary as the whole validation image to verify the overall accuracy and typical classification accuracy from global region. The accuracy information of the classification results was shown in Table 4. It could be seen from Table 4 that the accuracy of the SVM method based on linear, polynomial and RBF kernels, and the DCNN model combing the spectral feature and texture feature after K-L transformation were both higher than that of the DCNN model using the spectral feature only. Among all the methods that combined spectral feature with texture feature after K-L transformation, the classification accuracy of the DCNN model was the highest, the accuracy was 84.64%, and the Kappa coefficient was 0.80.

Among all the classification results of typical types, the developed DCNN model classification algorithm ensured the accuracy of all types more balanced, and it greatly improves the accuracy of tidal flat and farmland, while kept the classification accuracy of main types almost invariant compared to the shallow algorithms. The SVM-polynomial method that used in water had the highest recognition accuracy was 89.42% in the case of combined spectral feature and texture feature information after K-L

Table 3. The classification results (%) of testing samples

Type	Classification method														
	SVM-linear			SVM-polynomial			SVM-RBF			SVM-sigmoid			DCNN		
	SO	ST	STKL	SO	ST	STKL	SO	ST	STKL	SO	ST	STKL	SO	ST	STKL
Reed	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Tamarix	99.07	98.76	99.84	100	98.76	100	98.76	98.45	98.76	18.01	18.94	21.12	99.38	97.67	97.52
Spartina	95.71	95	99.64	93.57	94.29	97.8	92.85	94.29	97.5	21.07	20.71	51.78	98.93	95.35	96.79
Water	93.59	90.08	96.28	100	93.39	100	100	96.28	100	92.56	91.32	94.62	97.93	95.04	100
Tidal flat	95.95	94.72	98.41	96.48	94.01	98.24	98.24	95.42	98.77	84.51	91.55	90.14	98.42	96.94	98.59
Farmland	92.99	92.36	98.09	91.08	91.72	95.86	92.99	91.72	97.13	70.16	48.6	62.74	97.77	95.69	98.41
OCA	96.63	95.56	98.75	97.63	95.85	98.96	97.85	96.56	98.89	65.37	64.33	69.84	98.82	95.23	99.38

Note: SO represents spectral only, ST spectral + all texture, and STKL spectral + texture after K-L transform.

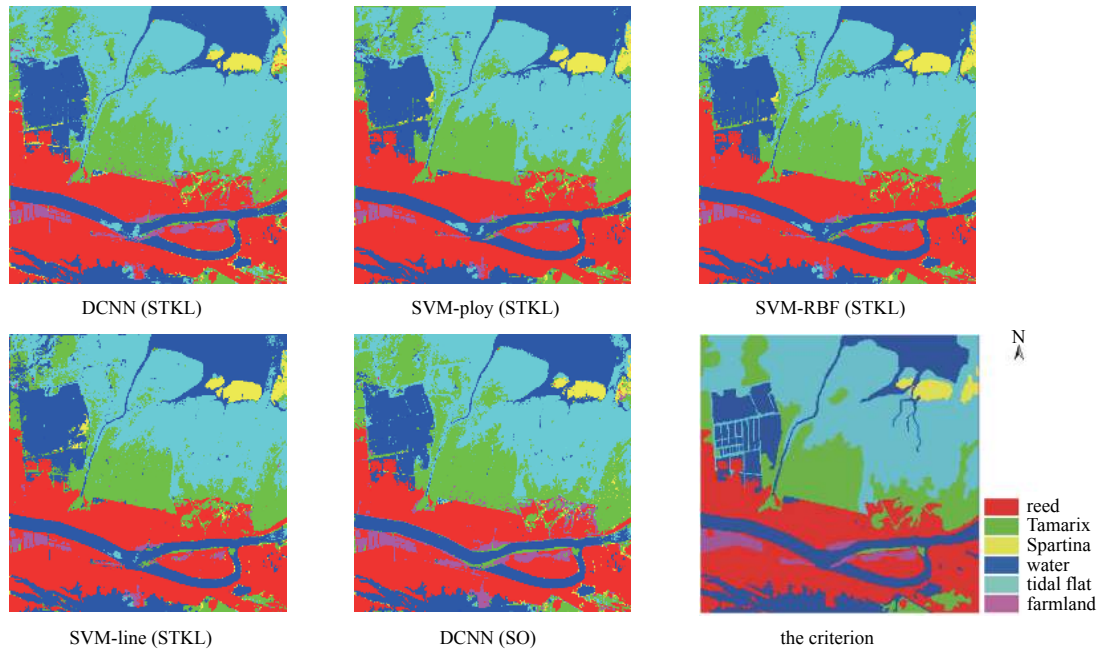


Fig. 6. The results of each classification method and the criterion.

Table 4. The classification results based on the whole validation image (%)

Type	Classification method														
	SVM-linear			SVM-polynomial			SVM-RBF			SVM-sigmoid			DCNN		
	SO	ST	STKL	SO	ST	STKL	SO	ST	STKL	SO	ST	STKL	SO	ST	STKL
Reed	94.2	96.61	95.35	92.75	96.78	94.94	96.06	96.99	94.63	88.89	89.36	91.4	92.57	92.99	94.01
Tamarix	77.71	76.52	79.91	80.17	78.07	83.62	78.99	78.76	82.46	19.2	19.88	27.66	78.33	78.54	79.24
Spartina	79.91	78.69	85.51	87.92	79.69	87.02	82.6	78.62	87.27	15.47	15.41	31.96	76.14	75.32	79.81
Water	83.32	78.11	88.67	86.5	78.59	89.42	83.63	80.64	89.39	69.75	68.39	73.01	85.31	85.23	86.36
Tidal flat	73.64	71.43	76.75	72.52	71.44	76.15	74.09	70.75	75.75	90.52	91.65	89.16	77.96	75.84	79.26
Farmland	43.99	45.91	38.78	45.34	45.42	43.81	42.1	45.34	45.18	46.12	31.13	35.15	51.85	52.37	56.72
OCA	81.04±	79.70±	83.80±	81.75±	80.11±	84.43±	81.95±	80.48±	84.06±	71.57±	71.46±	73.84±	82.68±	82.14±	84.64±
Kappa	2.26	0.88	0.12	1.01	0.13	0.04	1.60	0.65	0.07	1.06	0.95	0.41	0.15	0.39	0.20
	0.75	0.73	0.78	0.76	0.74	0.79	0.76	0.74	0.71	0.61	0.61	0.64	0.77	0.77	0.8

transformation. Tidal flat is the most widely covered area in the coastal wetland of Huanghe River Estuary. The most two difficult types to classify in the coastal wetland are tidal flat and farmland. Compared with other classification methods, the developed DCNN model can effectively extract them, and the extraction accuracy can be up to 79.26% and 56.72%, respectively. While the accuracy of them by the SVM methods with linear, polynomial and RBF kernel was less than 76.75% and 46.12%, and the accuracy improving about 2.51% and 10.6%.

In each classification method, the classification accuracy of combined spectral feature and all the texture feature was lower than the accuracy based on spectral feature and combined spectral feature and texture feature after K-L transformation. In the SVM methods, the classification accuracy of combined spectral feature and texture feature after K-L transformation improved by about 1% and 4%. In the three different cases by DCNN model, we can see that the classification accuracy was over 82%, and the classification accuracy by using all texture feature is the lowest, it was 82.14%, the classification accuracy by using spectral feature and texture feature after K-L transformation was up to 84.64%, and it improved by 2.5% compared with using all the texture feature.

3.3 The DCNN model application

Considering the applicable area of the proposed model and the accuracy of training samples and testing samples, the CHRIS hyperspectral image of the Huanghe River Estuary coastal wetland was obtained in September 2009, whose operating mode stays at State 2 and image angle is 0°. The image scene has 210×262 pixels and 18 spectral bands. There are 4 classes in this data. We randomly chose 730 training samples and 370 testing samples. Based on the selected samples, we adapted the developed DCNN model and SVM method to classify the hyperspectral image.

The classification results based on different methods are shown in Fig. 7. The results showed that the classification accuracy of the DCNN model combined spectral feature and texture feature after K-L transformation was the highest, the accuracy was 90.05%. In the three different cases by DCNN model, we could see that the classification accuracy by using all texture feature was the lowest, it was 72.23%, and it reduced by 17.72% and 14.25% compared with using spectral only and combined spectral feature and texture feature after K-L transformation. The developed DCNN model can be applied to the coastal wetland hyperspectral images classification.

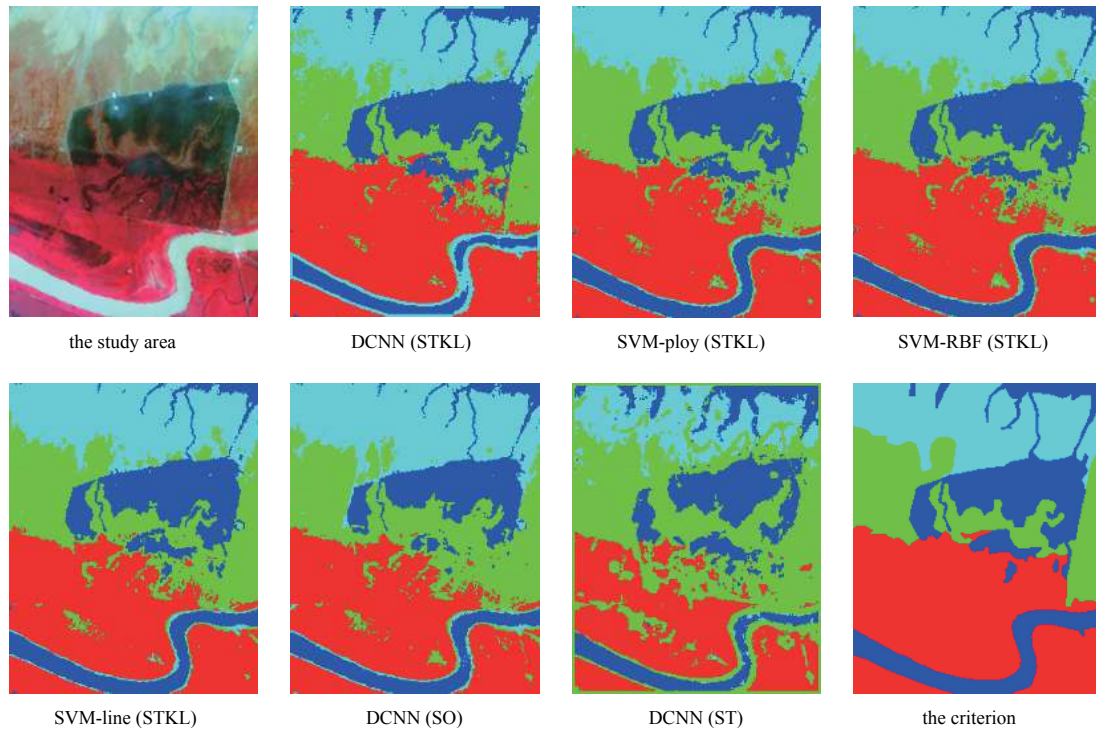


Fig. 7. The study area, the classification results and the criterion.

4 Conclusions

Coastal wetland locates in a transitional belt between land and marine system that carry out mutual material flow and energy transfer. It has important ecological and economic value. It is great significance to study the distribution and evolution of coastal wetland types. This paper developed a deep convolution neural network classification algorithm with fully-connected 8-layers based on CHRIS hyperspectral images, which combined spectral feature and multi-spatial texture feature information. We compared it with the SVM classification algorithm using the kernel of linear, polynomial, radial basis function and sigmoid function. At the same time, we evaluated the hyperspectral images classification accuracy with different input information. In the end, we validated the accuracy of the developed DCNN model for coastal wetland classification based on remote sensing interpretation map of coastal wetland of the Huanghe River Estuary. The results are shown as follows.

(1) Based on CHRIS hyperspectral remote sensing image of the coastal wetland, this paper constructed a classification system for typical object types of coastal wetland, and developed a fully-connected DCNN remote sensing classification model with 8-layer structure for coastal wetland, which combines spectral feature and multi-spatial texture information.

(2) From the verification of training sample, the developed DCNN classification algorithm model had high classification accuracy for tidal flat, and the classification accuracy reached up to 99%. The SVM classification algorithm based on sigmoid kernel had the lowest accuracy and the overall classification accuracy was less than 70%. The overall accuracy of testing sample was up to 99% based on the DCNN model combined spectral feature and texture feature after K-L transformation only. In each classification method, compared with the two cases that based on spectral feature and combined spectral feature and all the texture feature, the classification accuracy of combined spectral feature and tex-

ture feature after K-L transformation improves by about 2% and 3%. In the three different cases by DCNN model, we could see that the classification accuracy by using spectral feature and texture feature after K-L transformation was up to 99.38%, and it improved by 4.15% compared with using all the texture feature. Partial texture features would directly affect the classification accuracy.

(3) From the verification of whole image, the developed DCNN model classification method had the highest classification accuracy for coastal wetland of Huanghe River Estuary, which was 84.64%, and the Kappa coefficient was 0.80. Compared with other classification methods, the developed DCNN model can effectively extract them, with the accuracy of up to 79.26% and 56.72%, respectively. While the accuracy of them by the SVM methods with linear, polynomial and RBF kernel was less than 76.75% and 46.12%, improving about 2.51% and 10.6%. In the three different cases using DCNN model, we could see that the classification accuracy was over 82%, and the classification accuracy by using all texture feature is the lowest, it was 82.14%, the classification accuracy by using spectral feature and texture feature after K-L transformation was up to 84.64%, and it improved by 2.5% compared with that using all the texture feature.

(4) The developed DCNN model classification algorithm ensured the accuracy of all types more balanced. It also greatly improved the accuracy of tidal flat and farmland, while kept the classification accuracy of main types almost invariant compared to the shallow algorithms. Compared with other classification methods, the developed DCNN model performed better. The extraction accuracy can be up to 79.26% and 56.72%, respectively, and the accuracy improved about 2.51% and 10.6% compared with those by SVM methods with linear, polynomial and RBF kernel.

The coastal wetland classification algorithm based on the developed deep learning DCNN model was applied in the CHRIS

hyperspectral image of the Huanghe River Estuary, but whether it was suitable for other satellite imagery and other areas of coastal wetland classification was still to be further studied. On the other hand, there were some insufficiencies in the selection of the number of samples. The less selection samples of object types would affect the classification accuracy. Our future work will focus on solving the problem of sample selection.

References

- An Ni, Ma Yi, Bao Yuhai. 2016. Spectral fidelity analysis of scaling transformation of hyperspectral remote sensing image based on empirical mode decomposition. *Remote Sensing Technology and Application (in Chinese)*, 31(2): 230–238
- Cao Linlin, Li Haitao, Han Yanshun, et al. 2016. Application of convolutional neural networks in classification of high resolution remote sensing imagery. *Science of Surveying and Mapping (in Chinese)*, 41(9): 170–175
- Chen Yushi, Lin Zhouhan, Zhao Xing, et al. 2017. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6): 2094–2107
- Chubey M S, Franklin S E, Wulder M A. 2006. Object-based analysis of ikonos-2 imagery for extraction of forest inventory parameters. *Photogrammetric Engineering & Remote Sensing*, 72(4): 383–394
- Farabet C, Couprie C, Najman L, et al. 2013. Learning hierarchical features for scene labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8): 1915–1929, doi: [10.1109/TPAMI.2012.231](https://doi.org/10.1109/TPAMI.2012.231)
- Freund Y. 1995. Boosting a weak learning algorithm by majority. *Information and Computation*, 121(2): 256–285, doi: [10.1006/inco.1995.1136](https://doi.org/10.1006/inco.1995.1136)
- He Y, Qian D, Ben M. 2010. Decision fusion on supervised and unsupervised classifiers for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing Letters*, 7(4): 875–879
- Hinton G, Deng Li, Yu Dong, et al. 2012. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6): 82–97, doi: [10.1109/MSP.2012.2205597](https://doi.org/10.1109/MSP.2012.2205597)
- Hinton G E, Osindero S, Teh Y W. 2014. A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7): 1527–1554
- Hinton G E, Salakhutdinov R R. 2006. Reducing the dimensionality of data with neural networks. *Science*, 313(5786): 504–507, doi: [10.1126/science.1127647](https://doi.org/10.1126/science.1127647)
- Hu Wei, Huang Yangyu, Wei Li, et al. 2015. Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors*, 2015: 258619
- 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*. Lake Tahoe, Nevada: Curran Associates Inc, 1097–1105
- Lee H, Kwon H. 2016. Contextual deep CNN based hyperspectral classification. In: *Proceedings of 2016 IEEE International Geoscience and Remote Sensing Symposium*. Beijing, China: IEEE
- Li Wei, Prasad S, Fowler J E, et al. 2012. Locality-preserving dimensionality reduction and classification for hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 50(4): 1185–1198, doi: [10.1109/TGRS.2011.2165957](https://doi.org/10.1109/TGRS.2011.2165957)
- Li Xiaomin, Zhang Jie, Ma Yi, et al. 2015. Research on the classification method of the hyper-spectral image based on principal component analysis and decision level fusion. *Marine Sciences*, 39(2): 25–34
- Licciardi G, Pacifici F, Tuia D, et al. 2009. Decision fusion for the classification of hyperspectral data: outcome of the 2008 GRS-S data fusion contest. *IEEE Transactions on Geoscience and Remote Sensing*, 47(11): 3857–3865, doi: [10.1109/TGRS.2009.2029340](https://doi.org/10.1109/TGRS.2009.2029340)
- Mei Shaohui, Ji Jingyu, Bi Qianqian, et al. 2016. Integrating spectral and spatial information into deep convolutional Neural Networks for hyperspectral classification. In: *Proceedings of 2016 International Geoscience and Remote Sensing Symposium*. Beijing, China: IEEE, 5067–5070
- 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, doi: [10.1109/TGRS.2004.831865](https://doi.org/10.1109/TGRS.2004.831865)
- Slavkovicj V, Verstockt S, De Neve W, et al. 2015. Hyperspectral image classification with convolutional neural networks. In: *Proceedings of the 23rd ACM International Conference on Multimedia*. Brisbane, Australia: ACM, 1159–1162
- Sun Junjie, Ma Daxi, Ren Chunying, et al. 2013. Method of extraction of wetlands' information in nanweng river basin based on multi-temporal environment satellite images. *Wetland Science (in Chinese)*, 11(1): 60–67
- Tarabalka Y, Benediktsson J A, Chanussot J. 2009. Spectral-spatial classification of hyperspectral imagery based on partitioning clustering techniques. *IEEE Transactions on Geoscience and Remote Sensing*, 47(8): 2973–2987, doi: [10.1109/TGRS.2009.2016214](https://doi.org/10.1109/TGRS.2009.2016214)
- Teoh S S, Bräunl T. 2012. Symmetry-based monocular vehicle detection system. *Machine Vision and Applications*, 23(5): 831–842, doi: [10.1007/s00138-011-0355-7](https://doi.org/10.1007/s00138-011-0355-7)
- Tian Zhuangzhuang, Zhan Ronghui, Hu Jiemin, et al. 2016. SAR ATR based on convolutional neural network. *Journal of Radars (in Chinese)*, 5(3): 320–325
- Waibel A, Hanazawa T, Hinton G E, et al. 1989. Phoneme recognition using time-delay neural networks. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(3): 328–339, doi: [10.1109/29.21701](https://doi.org/10.1109/29.21701)
- Xu Yingxue, Shao Jingli, Yang Wenfeng, et al. 2006. Research on classification and change of seaside wetland around Yalujiang river estuary based on RS and GIS. *Geoscience (in Chinese)*, 20(3): 500–504
- Xu Zhenlei, Yang Rui, Wang Xinchun, et al. 2016. Based on leaves convolutional neural network recognition algorithm. *Computer Knowledge and Technology (in Chinese)*, 12(10): 194–196
- Yang Jingxiang, Zhao Yongqiang, Chan J C W, et al. 2016. Hyperspectral image classification using two-channel deep convolutional neural network. In: *Proceedings of 2016 International Geoscience and Remote Sensing Symposium*. Beijing, China: IEEE, 5079–5082
- Yue Jun, Zhao Wenzhi, Mao Shanjuan, et al. 2015. Spectral-spatial classification of hyperspectral images using deep convolutional neural networks. *Remote Sensing Letters*, 6(6): 468–477, doi: [10.1080/2150704X.2015.1047045](https://doi.org/10.1080/2150704X.2015.1047045)