Accurate Classification of Coastal Wetland in Yellow River Delta based on ResNet combined with Coordinate Attention

Wenli Wu [1], Delong Kong [2], Moughal Tauqir [3*], Jiahua Zhang [4]
[1-4] Remote Sensing Information and Digital Earth Center, College of Computer Science and Technology, Qingdao University, Qingdao 266071, China

* Corresponding Author: **Moughal Tauqir**

Article Info

P-ISSN: 3051-3383
E-ISSN: 3051-3391
Volume: 06
Issue: 02
July - December 2025
Received: 05-09-2025
Accepted: 07-10-2025
Published: 06-11-2025
Page No: 64-75

Abstract
The Yellow River Delta (YRD) is a typical coastal wetland in China, dominated by coastal salt marsh vegetation, the YRD wetlands are continuously converted by inland sediment and impact of waves and storm surges, so the accurate classification of the coastal wetlands in the YRD is important for the development and protection of wetland resources. In this study, we adopted deep learning methods and integrated remote sensing information to accurately identify the coastal wetland types of the YRD. At first, to construct a classification system for considering the features of the YRD, and then the labelled data are annotated based on the Sentinel-2 satellite data by using the measured data as a reference basis. Secondly, the CNN, ResNet-50 and MobileNetV1 were selected for fine classification of coastal wetland. In order to improve the classification accuracy, this paper introduces the coordinate attention mechanism in the ResNet-50, and the new network named CAResNet-50 was proposed. The results show that, compared with ResNet-50, the kappa coefficient of the classification result of CAResNet-50 is improved by 0.03, and the overall accuracy is increased by 2.68%. After the comprehensive analysis of several evaluation indexes, the CAResNet-50 model has the best classification performance at the coastal wetland classification in the YRD. The results indicated that the CAResNet-50 model, which introduces the coordinate attention mechanism, provides a new method for the accurate classification of coastal wetland in the region scales.

DOI: https://doi.org/10.54660/IJAIET.2025.6.2.64-75

Keywords: Deep Learning, Wetland Classification, Coastal Region, Remote Sensing

1. Introduction

The coastal wetlands, as an important part of blue carbon ecosystems (Duarte *et al*., 2016) [5], have a higher carbon sequestration capacity than terrestrial ecosystems and contribute to slowing down global warming. Tidal marshes, mangroves and seagrasses are the three main blue carbon ecosystems on the Earth (Chastain *et al*., 2021; Wang *et al*., 2021) [2, 31].The Yellow River Delta in China is a typical tidal marsh ecosystem with a large variety of marsh vegetation, such as reeds, suaeda, and tamarisk, which is an important habitat for endangered wild birds and a transit point for a large number of migrating birds.

The accurate classification of coastal wetland features has always been an important foundation for studying on coastal wetlands on basis of remote sensing image interpretation. The coastal wetland ecosystems are characterized by structural complexity, high degree of diversity and regional differences, and from the perspective of different purposes and disciplines, most of the researches adopted different feature classification systems to identify coastal wetland types, taking into account the special characteristics of different categories, in the wetland regions from remote sensing satellite data (Mcleod *et al*., 2011; Mou *et al*., 2015) [23, 25].

At present, the traditional methods for coastal wetland classification are unsupervised classification, visual interpretation and supervised classification. Along with the rapid development of automated extraction technology, support vector machine, decision tree method, object-oriented method, K-nearest neighbor algorithm and deep learning are applied to the research field of coastal wetland feature classification (Mohammadimanesh *et al*., 2015; Rasel *et al*.,2019; Li *et al*., 2024) [18, 24, 26].

Along with the rapid development of computer science and technology, the classification technology of coastal wetland based on remote sensing images is also developed, and machine learning algorithms have been widely used in wetland feature classification research. In order to differentiate between a variety of wetland types in the Avalon Peninsula of the Lambrado, Canada, and Newfoundland, Mohammadimanesh *et al*. (2018) [24] investigated a random forest algorithm based on hierarchical objects to achieve overall accuracy of the algorithm reached 94.82%. Zun *et al*. (2015) [35] used Nanjing wetlands as the study area, used high spatial resolution remote sensing image data, and classified the wetlands using an artificial neural network that processes information in a way similar to the synaptic connection structure of the brain, which improved the classification accuracy compared with the maximum likelihood classification algorithm. Rasel *et al*. (2019) [26] explored the relationship between spectral separability, kernel smoothing parameter and support vector machine for limited samples, and found an optimized support vector machine classification algorithm superior to the predecessors in wetland feature classification. (Salas *et al*., 2024) [29] used Sentinel-2 remotely sensed image data to classify multiple small wetlands in Beavercreek, Ohio, USA, and based on machine learning algorithms including classification and regression trees, random forests, K-nearest neighbor algorithms, plain Bayes, gradient tree augmentation, and support vector machines, and the results of the experiments found that gradient tree augmentation and random forests algorithms outperformed other classification algorithms.

Comparing with machine learning approach, deep learning can fully mine the underlying features of an image and learn the internal mechanism of these features. Therefore, in the context of the rapid development of computers, deep learning has been successfully applied in the field of image classification and significant results have been achieved. The remote sensing hyperspectral images satisfy the requirements of deep learning in terms of sample information, and deep learning models have received widespread attention in the field of hyperspectral image classification in recent years(Cao *et al*., 2018; Li *et al*., 2019; Santara *et al*., 2017; Yang *et al*., 2018) [1, 16, 30, 32]. Fang *et al*. (2020) [6] proposed a novel collaborative learning framework for semi-supervised hyperspectral image classification with joint deep convolutional neural networks (CNNs) and deep clustering, to solve the problem of limited labeled training samples in the hyperspectral image datasets. Considering that transfer learning strategies have the potential for the hyperspectral image classification, He *et al*. (2020) [9] proposed a new classification framework that combined transfer learning and deep CNN for hyperspectral image classification, and achieved good results especially when the training samples are limited. To fully take advantage of spatial and spectral information of the hyperspectral image, Han *et al*. (2020) [8] introduced a new joint spatial–spectral hyperspectral image classification method based on different-scale two-stream convolutional network and spatial enhancement strategy. Roy *et al*. (2020) [27] developed a hybrid spectral CNN (HybridSN) for hyperspectral image classification, which is a spectral–spatial 3-D CNN followed by spatial 2-D-CNN. HybridSN model combines the complementary information of spatial-spectral and spectral in the form of 3-D and 2-D convolutions, respectively. Deep belief networks (DBNs) combine unsupervised learning and supervised learning, and it was firstly proposed by Hinton *et al*. (2006) [10] in 2006, the DBNs use an unsupervised process to train the network parameters of each layer, and then apply a supervised process to calculate and transfer errors, constantly updating the network parameters through some optimization method. Finally, the network model can effectively describe the characteristics of the data and implement classification. Many reports also show that the DBN classification model can be successfully applied to hyperspectral classification (Tong *et al*., 2017; Li *et al*., 2018; P. Liu *et al*., 2017) [7, 15, 20].

Under the influence of the rapid development of computer technology, deep learning is also widely used in the classification of coastal wetlands. Li *et al*. (2021) [14] selected Sentinel-2 bands and normalized exponential indices, utilized U-net and proposed an adaptive deep learning approach to map the emergent wetlands of the estuary of South Carolina, USA. and the overall classification accuracy of the final experimental results was 90%. Luo *et al*. (2024) [21] used local and global features to improve the classification accuracy of coastal wetlands, and proposed a dual-scale convolutional self-attention network consisting of a dual-scale summation module and a convolutional self-attention module, and the overall accuracies of the experimental results on two hyperspectral datasets from Zhuhai-1 satellite were 98.29% and 96.82%, respectively. Lin *et al*. (2023) [19] developed the SegFormer model based on Sentinel-2 remote sensing maps to classify 9 types of features in coastal wetlands of Yancheng City, Jiangsu Province, and obtained good classification results with an overall accuracy of 94%. Li *et al*. (2023) [17] quantified the spatial and temporal changes of wetlands on the Tibetan Plateau from 1990 to 2019 based on Landsat imagery using the SE-ResNet-50 model, and the final wetland extraction had an overall accuracy of 95.5%.

The complex feature types and distributions of the remote sensing images of the coastal wetlands in the YRD bring challenges to the accurate classification, but it is expected to obtain a high-precision classification of the complex features of the coastal wetlands through deep learning to classify the remote sensing images of the study area on a pixel-by-pixel basis. The main objectives of this study are: to establish the classification system of coastal wetland features in the YRD and constructing the dataset, to introduce the coordinate attention mechanism in ResNet-50 and propose the new model CAResNet-50, and adopt the CNN, ResNet-50, MobileNetV1 and CAResNet-50 methods to perform fine classification based on the Sentinel- 2 multispectral satellite remote sensing data for the fine classification of coastal wetlands in the YRD region.

2. Study area and data

2.1. Study area

The Yellow River Delta (YRD) is located at the mouth of the Yellow River in Shandong Province, China, between longitude 117°31′ and 120°32′ E and latitude 36°55′ and 38°16′ N. It covers an area of 12,038 square kilometers (km^2) (Ma *et al*., 2019) [22] (Figure. 1). The YRD has a continental temperate monsoon climate with four distinct seasons and an average annual rainfall of 530-560 mm, with summer rainfall accounting for about 70% of the year (Yu *et al*., 2012) [33]. The summer rainfall accounts for about 70% of the year.

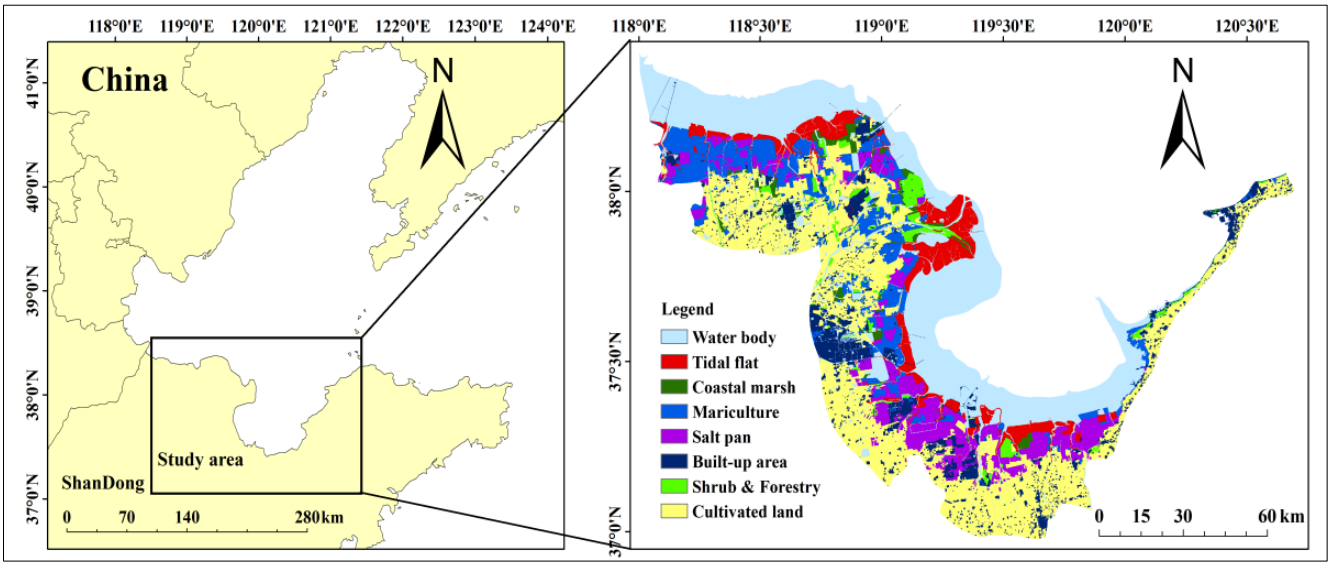

Fig 1: Location map of the Yellow River Delta study area and its 2015 LULC types

The YRD wetland is roosting and feeding place of birds in every year, making it as a transit point for bird migration. When the upper reaches of the Yellow River enter the sea, a large amount of sediment carried by the river is deposited in the downstream depression to form the YRD, which is the second largest estuarine delta in China (Zhao *et al*., 2022) [34].

The alluvial deposits of the Yellow River form the main part of the soil matrix of the YRD, and tidal soil and salt soil are the main soil types. The estuary of the YRD is a typical coastal wetland blue carbon ecosystem in China, dominated by coastal salt marsh vegetation including reeds, suaeda, tamarisk, and spartina alterniflora etc. (Chen *et al*., 2022) [3].

Table 1: Sentinel-2A satellite image parameters

waveband number	Spatial resolution (m)	Bandwidth (nm)	Wavelength (nm)
B1-Coastal /Aerosols	60	27	433-453
B2-Blue	10	98	458-523
B3-Green	10	45	543-578
B4-Red	10	38	650-680
B5-Vegetation Red Edge	20	19	698-713
B6-Vegetation red edge	20	18	733-748
B7-Vegetation Red Edge	20	28	773-793
B8-Near infrared	10	145	785-900
B8a-Red edge of vegetation	20	33	848-881
B9-Water vapour	60	26	935-955
B10-Curling clouds	60	75	1360-1390
B11-Shortwave infrared	20	143	1565-1655
B12-Shortwave infrared	20	242	2100-2280

2.2 Data

Sentinel-2 is a high-resolution multispectral satellite with a revisit period of 10 days and a multispectral imager, consisting of two satellites, Sentinel-2A and Sentinel-2B. The two satellites are complementary, with a revisit period of 5 days (Drusch *et al*., 2012) [4]. Sentinel-2A and Sentinel-2B were launched in 2015 and 2017, respectively. When both 2A and 2B are in the operational state, the period to form a complete image at higher latitudes is three days, while around the equator the process takes five days. The two satellites have an average altitude of 786 km and a swath width of 290 km, they are synchronized with each other in a 180 km phase in a sun-synchronous orbit, and their positions in orbit are detected by dual-frequency GNSS receivers. In this study, the Sentinel-2A satellite images were used, and the main parameters of the satellite are shown in Table 1. The Sentinel-2A satellite covers a total of 13 bands, including the red-edge bands (B5, B6, B7, B8a), which can be used for vegetation monitoring. Bands 10, 11 and 12 belong to the short-wave infrared band. Bands 2, 3 and 4 are in the visible band, and the uses of Sentinel-2 data include the land environment, such as coastal zones, terrestrial vegetation and soil cover, as well as the monitoring of natural hazards such as landslides, floods and volcanic eruptions.

Sentinel-2 multispectral data are downloaded from the ESA Data Centre (https://scihub.copernicus.eu/). Sentinel L1C data are an atmospheric apparent reflectance data product that is geometrically fine-corrected and orthorectified at the sub-image level. Sentinel L2A is an atmospherically corrected Sentinel L2A is an atmospherically corrected atmospheric bottom reflectance data product. The Sentinel L1C product was pre-processed to produce the L2A product for feature classification. The processing software used is Sen2cor, officially released by ESA, and the Sentinel-2A data preprocessing steps involved in this study are mainly atmospheric correction and data resampling. The atmospheric correction of the downloaded Sentinel L1C data products was performed using the Sen2cor processing tool to obtain the L2A data products required for the classification of this paper. The 10th band of the Sentinel-2A data was excluded from the thin cirrus cloud, and a total of 12 bands were included in the obtained L2A products. The spatial resolution of the Sentinel-2A data products varied in each band, divided into 10m, 20m, 20m and 20m. The Sentinel-2A data product has different spatial resolutions for each band, 10 m, 20 m and 60 m. The L2A data with atmospheric corrections were resampled using the SNAP tool to achieve a spatial resolution of 10 m for each band.

Based on the previous report (Chen *et al*., 2022) [3], in this study, the coastal wetland in the YRD was classified into 10 types of estuarine features, namely, reeds, suaeda glauca, spartina alterniflora, mixed tamarisk, exposed tidal flat, water body, puddle and pond, robinia pseudoacacia forest, cultivated land, bare land.

3. Experimental methodology

As an important blue carbon ecosystem, the vegetation type of coastal wetlands in the YRD is dominated by marsh vegetation. This study focuses on fine classification of coastal wetland based on Sentinel-2 satellite images using the CNN, the residual network ResNet-50 and the lightweight network MobileNetV1. In order to improve the classification performance, a coordinate attention mechanism that can focus on both feature map channel information and spatial information is introduced in ResNet-50, and the new proposed model is named CAResNet-50.

3.1. Convolutional Neural Networks (CNN)

The CNN is a typical model in deep learning method. CNNs can automatically extract image abstract features when processing image data, and then mine the potential semantic features in the image. CNNs have been effectively applied in image processing fields such as semantic segmentation, image enhancement, and target detection. In last decade, the emergence of backpropagation algorithms has driven the development of CNNs the first well-established CNN to appear was the LeNet model, which was applied to recognize handwritten characters (Lecun *et al*., 1998; LeCun and Bengio, 1998; Rumelhart and McClelland, 1987) [12, 13, 28]. The structure of the CNN model in this paper contains three convolutional layers, two maximum pooling layers, two ReLu (Rectified Linear Unit) activation functions, a fully connected layer, and an output layer (Figure. 2).

The model training process is divided into forward propagation process and back propagation process. The forward propagation process of the model means that the original image pixels in the form of a two-dimensional matrix, through the convolution layer to obtain the initial feature map of the image, and the feature map is activated in the activation function. Then it is input into the pooling layer for the dimensionality reduction and to strengthen the invariance of the image features. The pooling layer is generally divided into the average pooling and the maximum pooling. Finally, through the fully connected layer to get the feature map to one-dimensional vector representation, and through the activation function to get the category of the image. The backpropagation process of the model means that the prediction error between the predicted values of the model and the real values of the data is measured by an objective function, which is used to guide the updating of the parameters of the model.

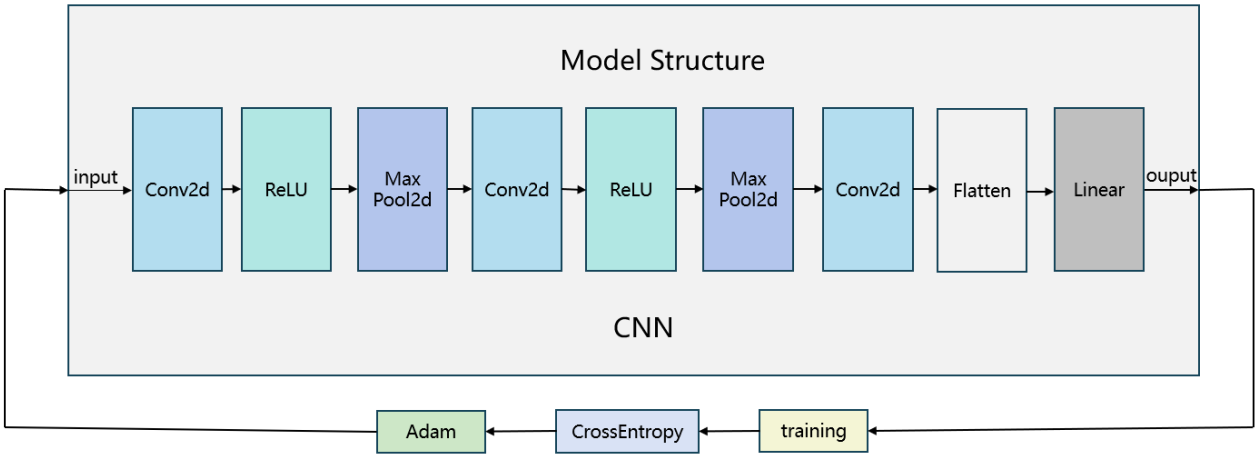

Fig 2: CNN model structure

Table 2: ResNet-50 network structure

Layer name	50-layer	output
Conv1	7×7, 64, stride 2, Batch Normal	112 x 112
	ReLu, Max pool, 3 x 3, stride 2	
Conv2_x	1 x 1, 64 3 × 3, 64 × 3 1 x 1, 256	56 x 56
Conv3_x	1 x 1, 128 3 x 3, 128 x 4 1 x 1, 512	28 x 28
Conv4_x	1 x 1, 256 3 × 3, 256 × 6 1 x 1, 1024	14×14
Conv5_x	1 x 1, 512 3 × 3, 512 × 3 1×1, 2048	7 × 7
	Average pool, FC, Softmax	1 × 1

3.2. The residual network ResNet-50

Along with the depth of the CNN continues to increase, the problem of gradient explosion or gradient disappearance will appear, there will be difficulties in convergence of the network. ResNet, the residual network, emerged as a solution to the above problems. Nowadays residual networks are widely used in computer vision related fields including video action recognition, image classification and image generation. According to the number of convolutional layers, the ResNet can be divided into different models such as ResNet-18, ResNet-50 and ResNet-152.

3.2.1. General network structure

The ResNet-50 network has one fully connected layer and 49 convolution layers (Table 2). The ResNet-50 can be divided into six parts. The multispectral satellite images are first processed by the first part of the model, which contains a convolutional layer, a regularization, a ReLu function and a maximum pooling. The size of the convolutional kernel is 7 × 7 × 64 and the number of channels of its output feature map is 64. The second to the fifth parts each contain a residual unit, which contains three convolution layers. The second to the fifth part of ResNet-50, each part is repeated 3 times, 4 times, 6 times, 3 times, so that there is a total of 1 + 3 × (3 + 4 + 6 + 3) = 49 convolution layers, plus one fully connected layer, a total of 50 layers, which is where the name of the ResNet-50 neural network comes from. The last part of ResNet-50 goes through an average pooling layer, a fully connected layer, and finally a Softmax classifier to get the probability of the class that the input data.

3.2.2. CAResNet-50 with the coordinate attention

The Coordinate Attention (CA) mechanism focuses on the channel and spatial information of the feature. Attention machines such as CBAM and SE usually use average pooling or maximum pooling when calculating the channel attention, which can discard the spatial information of the object [36]. Coordinate attention mechanisms also compute spatial attention when computing channel attention, introducing position information into the channel attention (Figure 3). The channel information refers to the information of the feature map on different channels, i.e., the difference and correlation between different features.; The spatial information refers to the positional relationship between different pixel points, i.e., the spatial structure and distribution of the pixel points. The CA mechanism pays attention to the channel information and spatial information of the feature layer implies that it considers the positional relationship and correlation between features at the same time in order to improve the model's feature extraction and characterization capabilities.

The CA mechanism starts with average pooling of size C × H × W in parallel in both the width and height directions, and the output feature maps of size C × H × 1 and C × 1 × W are obtained in the width and height directions, which map the features to the wide dimension and the high latitude, respectively. The two feature maps after average pooling are then merged to combine the high and wide feature information together. At this point, the feature map size is C/r×1×(H+W) and is processed for convolution, normalization and non-linear activation, with channel scaling

for convolution. r notes the scaling factor. After that, the feature is splited again in the wide and high dimensions for parallel operation, respectively Finally, the combined attention features are obtained through Sigmoid activation.

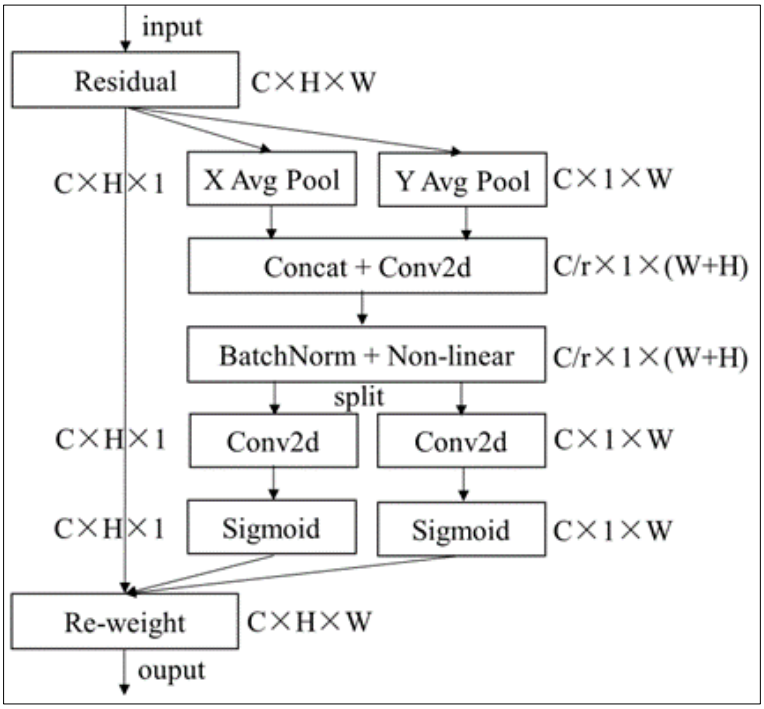

Fig 3: Structure of the coordinate attention mechanism

In neural network models, the numerous information of input feature that is prone to the problem of information overload. Introducing the attention mechanism in neural networks can increase the attention to critical information, reduce the connection to other irrelevant information, and make effective use of computational resources. In order to improve the classification accuracy and efficiency of remote sensing images, this paper introduces the coordinate attention mechanism into the ResNet-50 network and names the new network CAResNet-50, Figure 4 shows the structure of the CAResNet-50 network. CAResNet-50 performs a coordinate attention calculation after each of the first five parts of ResNet-50. Each coordinate attention mechanism calculation takes the output feature map of the previous section as the input feature map, and takes the output feature map of the calculation as the input feature map of the next section.

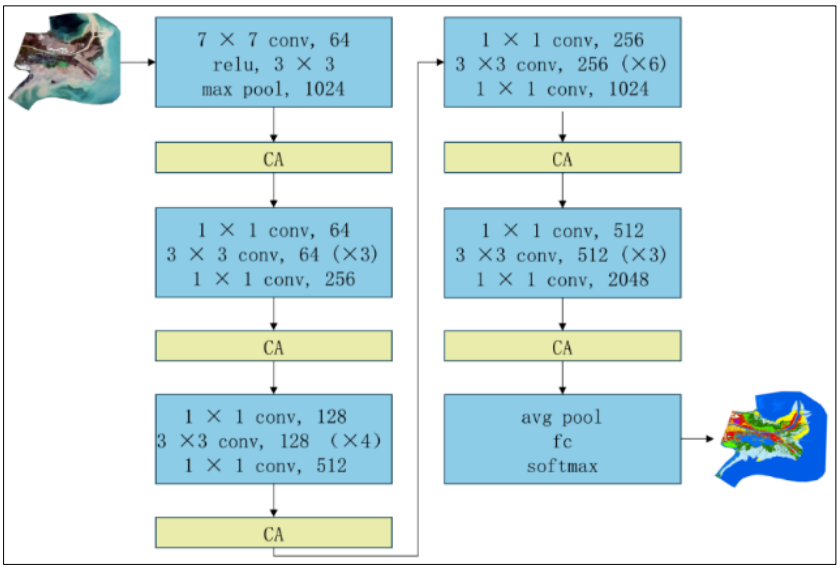

Fig 4: CAResNet-50 network structure diagram

3.3. The MobileNetV1 network

Traditional CNNs tend to be more computational and memory demanding. MobileNetV1 network is proposed for use on embedded devices or mobile. Depthwise separable convolution is the core of the network.

3.3.1. Depthwise separable convolution

The standard convolution operation simultaneously filters and combines input features into a new set of output features in a single step. In contrast, the depthwise separable convolution factorizes this process into two distinct layers: a

depthwise convolution followed by a pointwise convolution (Howard *et al.*, 2017) [11]. DW convolution and PW convolution. The DW convolution shown in Figure. 5 and then PW convolution shown in Figure. 6. First, the DW convolution applies a single convolutional filter to each input channel independently. It operates on the spatial dimensions (height and width) of each channel separately to extract features, but it does not combine information across channels. The number of output channels from this layer remains identical to the number of input channels. Next, the PW convolution follows. This is a simple 1x1 convolution whose role is to combine the output channels from the DW convolution. It projects these channels into a new, higher-dimensional feature space by computing linear combinations of them, thereby creating the desired number of output features.

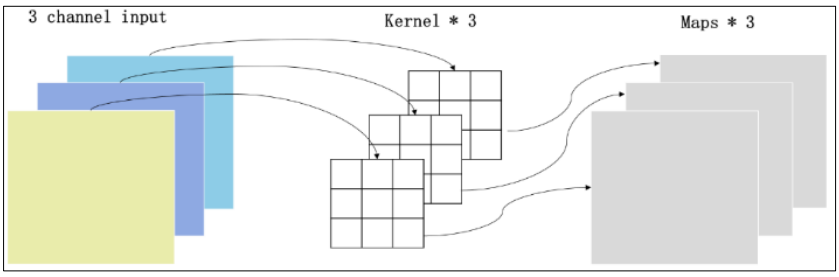

Fig 5: DW convolution in depth separable convolution

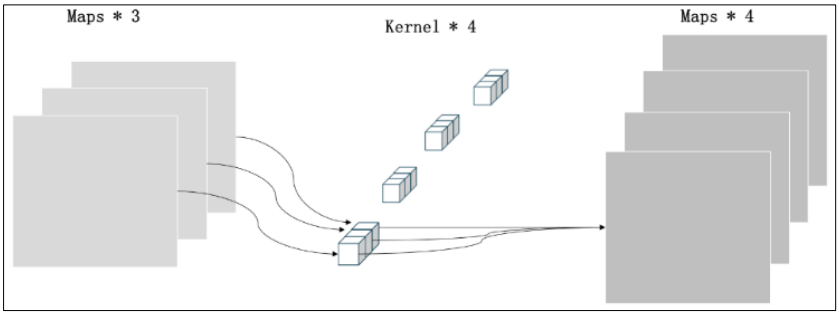

Fig 6: PW convolution in deep separable convolution

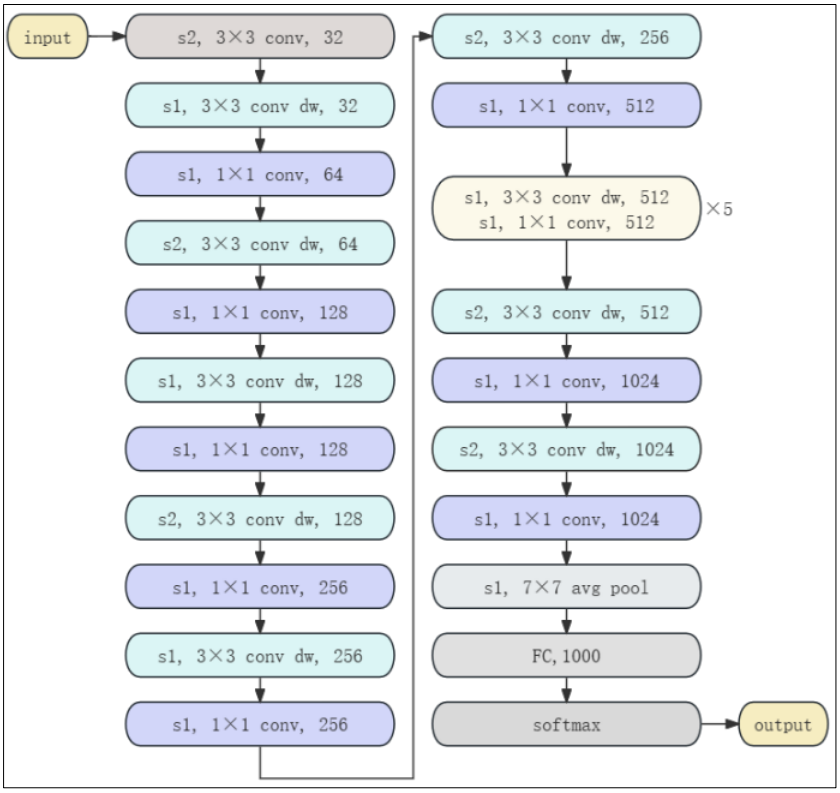

Fig 7: General structure of MobileNetV1 network

3.3.2 General network structure

Figure. 7 shows the structure of the MobileNetV1. s1 and s2 denote convolution stride of 1 and 2, respectively, conv denotes ordinary convolution, and conv dw denotes DW convolution. The MobileNetV1 firstly performs one ordinary convolution of the input image with a convolution kernel size

of 3 × 3, a channels number of 64, and a stride size of 2. Then it is to perform a depth-separable convolution, i.e., a DW convolution followed immediately by a PW convolution. After 13 depth-separable convolutions, the output is obtained through the fully-connected layer and the sigmoid function, followed by an average-pooling dimensionality reduction process. The MobileNetV1 totally undergoes 23 convolutional layers.

3.4. Criteria for evaluating classification results

In this paper, several evaluation metrics such as overall accuracy, precision, recall, and kappa coefficient are used to evaluate the model classification results. Overall accuracy refers to the ratio of the number of correctly predicted samples to the number of all samples. Precision refers to the proportion of correctly predicted samples to the predicted samples number in a certain prediction category. Recall is the proportion of the number of correctly predicted samples in a true category to the number of all samples in that category. The kappa coefficient is an indication of how well the two annotations fit together, where the annotation is the division of N samples into M mutually exclusive categories, and the closer the kappa value is to 1, the better the performance of the model.

$$overall\ accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$precision = \frac{TP}{TP + FP} \quad (2)$$

$$recall = \frac{TP}{TP + FN} \quad (3)$$

$$kappa = \frac{p_o - p_e}{1 - p_e} \quad (4)$$

The definition of each parameter in equation (1) is shown in Table 3. In equation (4), p_o represents the proportion of predicted and true values that match, and p_e notes the proportion of all matches generated by randomness.

Table 3: Meaning of parameters of the overall accuracy equation

		Projected results	
		positive	**negative**
real value	positive	True positive (TP)	False negative (FN)
	negative	False positive (FP)	True negative (TN)

Table 4: Model experiment parameter settings

Batch size	learning rate	Test set ratio	Training set ratio
512	0.001	30 per cent	70 per cent

Table 5: Comparison of the overall evaluation of the classification results with the models

Evaluation indicators	CNN	ResNet-50	MobileNetV1	CAResNet-50
Kappa	0.9318	0.9262	0.9275	0.9562
AA (%)	93.89	93.40	93.52	96.08

4. Results

4.1. Setting of Classified Data and Network Parameters

The Sentinel-2 satellite multispectral data of 29 September 2019 were downloaded from the ESA official website, and the estuary of the coastal wetland of the Yellow River Delta was obtained as the image classification study area after mask extraction, which contained 10,668,320 pixels. A total of 10 types of feature types of the YRD coastal wetland inlet were classified in the study area, which were saline alkali ponts, nurse miscanthus, reeds, potholes, waters, bare tidal flats, acacia forests, tamarisk mixes, harvested arable land or bare land, and arable land. A total of 84953 pixels were labelled as label samples for training and testing of convolutional neural network models. The number of batch training samples was 512 and the learning rate was 0.001 in each model of CNN, ResNet-50, MobileNetV1 and CAResNet-50 (Table 4). A total of 84953-pixel points of labelled samples were produced at the Yellow River Delta estuary in the study area, and 70% of the labelled samples were used as the training set and 30% of the samples were used as the test set for each category.

4.2. Coastal wetland classification results

In order to select the most suitable model among the CNN, ResNet-50, MobileNetV1 and CAResNet-50, for the classification of coastal wetlands in the YRD, this work presents a comprehensive comparative analysis of the feature accurate classification results of the four models in terms of the overall and each feature category. The Kappa coefficients of classification results of the four models CNN, ResNet-50, MobileNetV1 and CAResNet-50 are 0.9318, 0.9262, 0.9275 and 0.9562, and the overall accuracies are 93.89%, 93.40%, 93.52% and 96.08%, respectively (Table 5). From the overall evaluation, the CAResNet-50 model has the highest Kappa coefficient and overall accuracy. Compared to the ResNet-50 model, the Kappa coefficient of the CAResNet-50 model increased by 0.03 and the overall accuracy increased by 2.68%.

Table 6: Comparison of the classification results of each category with different models

Category	CNN		ResNet-50		MobileNetV1		CAResNet-50	
	P (%)	R (%)	P (%)	R (%)	P (%)	R (%)	P (%)	R (%)
Reeds	89.79	85.67	98.12	98.91	98.54	93.61	92.60	94.06
Suaeda glauca	94.14	87.94	72.58	99.78	99.97	95.19	98.19	96.51
Spartina alterniflora	96.22	92.74	100.00	99.28	72.14	100.00	82.96	100.00
Mixed tamarisk	89.11	94.08	98.01	62.27	94.66	97.07	98.63	98.10
Exposed tidal flat	97.37	96.89	99.76	97.24	99.97	99.74	97.67	100.00
Water body	97.28	99.16	88.10	100.00	99.91	100.00	99.16	99.94
Puddle and pond	99.17	97.60	99.92	84.69	100.00	99.90	100.00	98.89
Robinia pseudoacacia forest	86.18	99.18	98.40	99.87	100.00	73.68	100.00	82.36
Cultivated land	90.71	92.33	99.66	96.50	89.63	69.82	99.13	85.72
Bare land	96.83	97.90	95.31	98.16	94.89	100.00	99.94	100.00

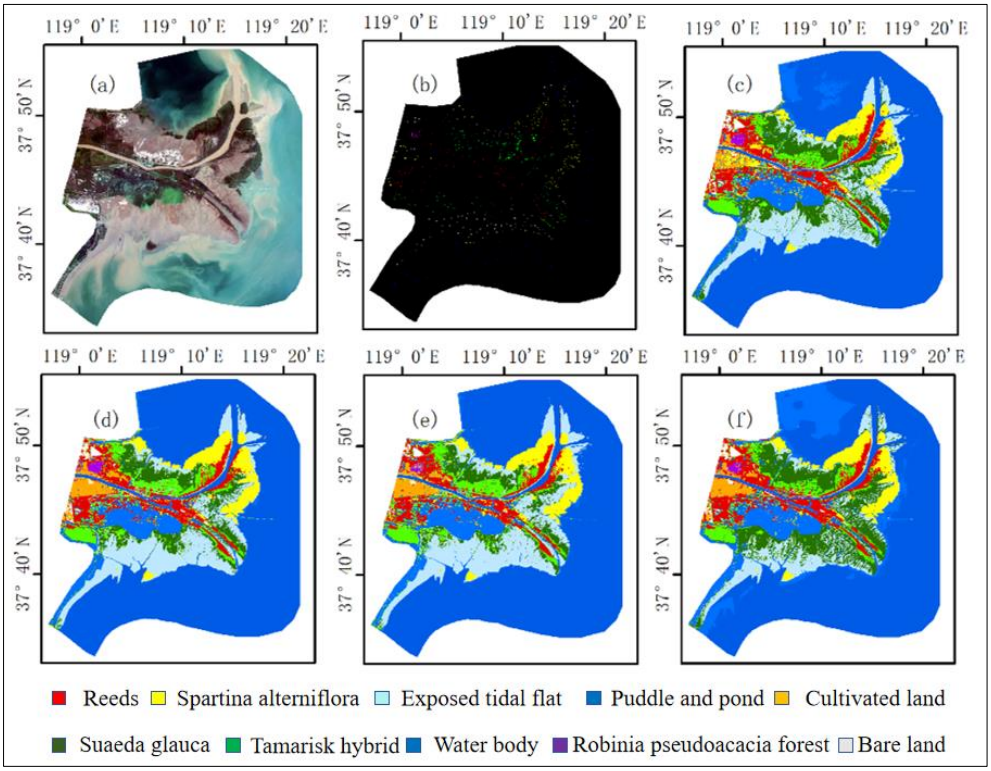

Fig 8: Classification map of the study area of the Yellow River estuary in the coastal wetlands of the Yellow River Delta. (a) Sentinel-2 satellite multispectral image map. (b) Labelling map. (c) CNN classification map. (d) ResNet-50 classification map. (e) MobileNetV1 classification map. (f) CAResNet-50 classification map.

In terms of accuracy, the CAResNet-50 model had classification accuracies higher than 97.50% for all categories except for reeds 92.60% and spartina alterniflora 82.96% (Table 6). Compared with the ResNet-50 model, the CAResNet-50 model had higher precision rates in seven categories, and the ResNet-50 model had one category with too low classification precision rate, i.e., suaeda glauca at 72.58%. In terms of recall, the CAResNet-50 model had a higher classification recall than 94% for all categories except for robinia pseudoacacia forest at 82.36% and cultivated land at 85.72%. The CAResNet-50 model had a recall of 1 for spartina alterniflora, exposed tidal flat, and bare land, suggesting that the model predicted all real samples correctly for these three categories. The MobileNetV1 model had a lower recall of 73.68% and 69.82% in robinia pseudoacacia forest and cultivated land, respectively. The ResNet-50 model had a lower recall of 62.27% in mixed tamarisk. From the precision and recall of each category together, the classification accuracy of CAResNet-50 is better than that of other three models.

The classification results of CNN, ResNet-50, MobileNetV1 and CAResNet-50 on Sentinel-2 satellite multispectral images are shown in Figure 8. The classification results of the CAResNet-50 model in the image are the same as the distribution of the actual data collected in the existing field, and combining with Tables 5 and 6, it shows that ResNet-50 can well identify different categories.

The confusion matrix from Figure 8 shows that, the most serious confusion in the classification results of the ResNet-50 model is the tamarisk hybrid, which has 1130 real samples misclassified as suaeda glauca. The misclassification of tamarisk hybrids into Suaeda glauca may be due to the fact that in some locations where soil salinisation is relatively mild, the suaeda glauca will grow taller and can be easily confused with tamarisk hybrid. In the CAResNet-50 model, the misclassification of tamarisk hybrid as suaeda glauca was improved. And compared with the confusion matrix of classification results of MobileNetV1 model, there were fewer samples of cultivated land being misclassified as spartina alterniflora. The overall evaluation of the

classification results of each model, the evaluation of the classification results of each category of each model and the confusion matrix of the classification results showed that CAResNet-50 had the best classification performance in the study area of the YRD compared with the other three models.

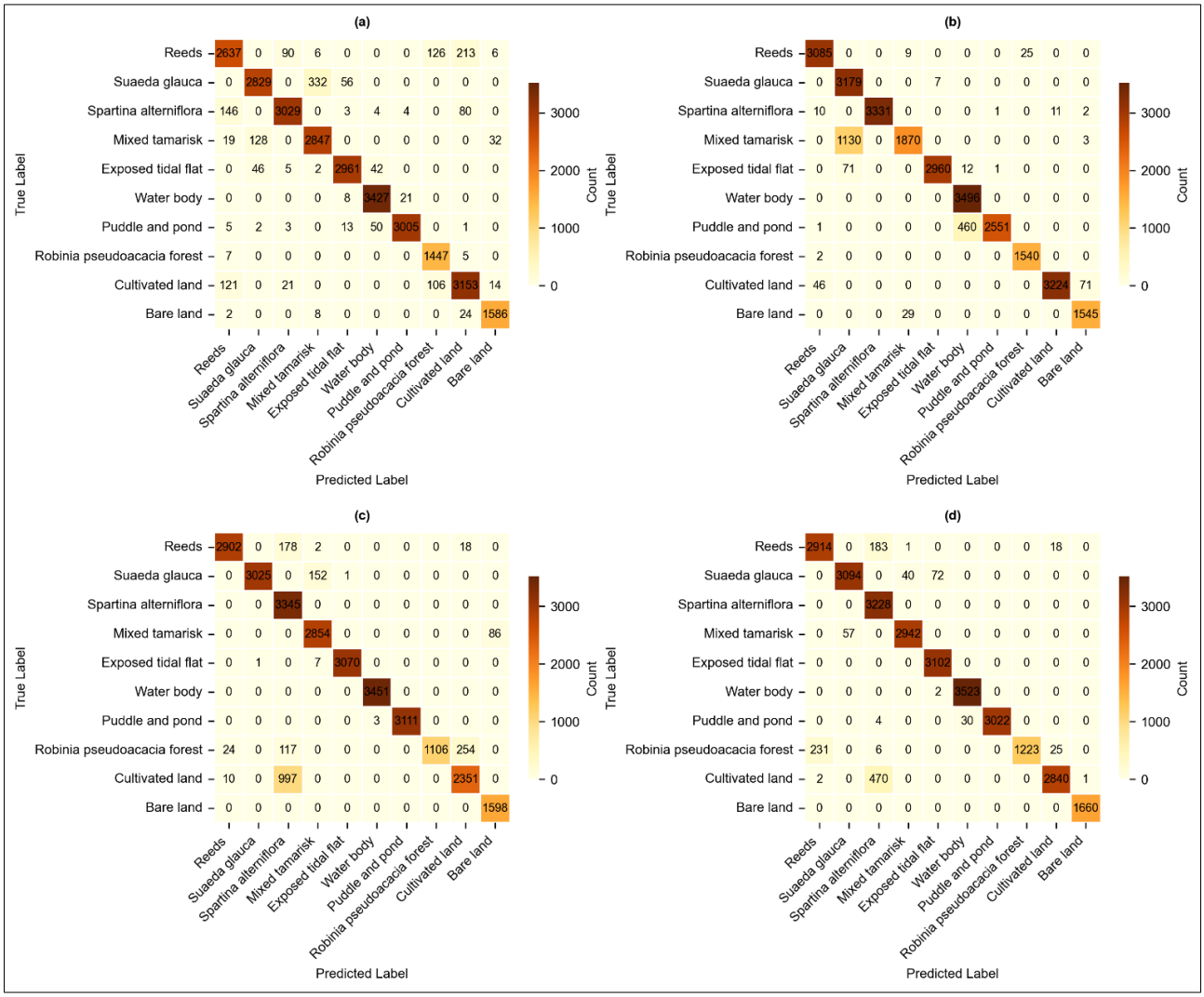

Fig 9: Confusion matrix of classification results for four models. (a) CNN, (b) ResNet-50, (c) MobileNetV1, and (d) CAResNet-50.

5. Conclusion

In this study, by using the super-resolution image of Sentinel-2 satellite, the accurate classification of coastal wetland in Yellow River Delta was obtained based on ResNet combined with coordinate attention. Firstly, we determined the scope of the study area and the feature classification system of the YRD. Then performed data preprocessing operations on Sentinel-2 remote sensing images, including atmospheric correction, resampling and cropping, and to label the remote sensing images with training samples needed for deep learning. Furthermore, we performed deep learning methods including CNN, ResNet-50, MobileNetV1 and CAResNet-50 integrated coordinate attention mechanism to classify wetland using Sentinel-2 remote sensing images. The classification results were used as evaluation criteria to comprehensively analyze the performance of each model via Kappa coefficient, overall accuracy, precision and recall. the results indicated the CAResNet-50 model has the best classification accuracy, and its classification results are taken as the final feature classification of coastal wetland in the Yellow River Delta.

Acknowledgments

This work was supported by Shandong Provincial Natural Science Foundation (No. ZR2020GF067), and the Discipline Cluster Research Project of Qingdao University: “Deep mining and intelligent prediction of multimodal big data for marine ecological disasters” (No. XT2024101).

Author Contributions

Wenli Wu: data curation, investigation, software, methodology, formal analysis, writing−original draft preparation. Delong Kong: formal analysis, investigation, validation, writing−review and editing. Moughal Tauqir: Funding acquisition, Project administration, writing−review and editing. Jiahua Zhang: conceptualization, supervision and writing−review and editing. All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have

appeared to influence the work reported in this paper.

6. References

1. Cao X, Zhou F, Xu L, Meng D, Xu Z, Paisley J. Hyperspectral Image Classification With Markov Random Fields and a Convolutional Neural Network. IEEE Trans Image Process. 2018;27:2354-67. doi: 10.1109/TIP.2018.2799324.
2. Chastain SG, Kohfeld KE, Pellatt MG, Olid C, Gailis M. Quantification of Blue Carbon in Salt Marshes of the Pacific Coast of Canada. Biogeosciences Discuss. 2021. doi: 10.5194/bg-2021-157.
3. Chen C, Ma Y, Ren G, Wang J. Aboveground biomass of salt-marsh vegetation in coastal wetlands: Sample expansion of in situ hyperspectral and Sentinel-2 data using a generative adversarial network. Remote Sens Environ. 2022;270:112885. doi: 10.1016/j.rse.2021.112885.
4. Drusch M, Del Bello U, Carlier S, Colin O, Fernandez V, Gascon F, *et al*. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Remote Sens Environ. 2012;120:25-36. doi: 10.1016/j.rse.2011.11.026.
5. Duarte CM, Losada IJ, Hendriks IE, Mazarrasa I, Mazarrasa NM. Correction: Corrigendum: The role of coastal plant communities for climate change mitigation and adaptation. Nat Clim Chang. 2016;6:802. doi: 10.1038/nclimate3062.
6. Fang B, Li Y, Zhang H, Chan JC-W. Collaborative learning of lightweight convolutional neural network and deep clustering for hyperspectral image semi-supervised classification with limited training samples. ISPRS J Photogramm Remote Sens. 2020;161:164-78. doi: 10.1016/j.isprsjprs.2020.01.015.
7. Guofeng T, Yong L, Lihao C, Chen J. A DBN for hyperspectral remote sensing image classification. In: 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA); 2017. p. 1757-62. doi: 10.1109/ICIEA.2017.8283123.
8. Han M, Cong R, Li X, Fu H, Lei J. Joint spatial-spectral hyperspectral image classification based on convolutional neural network. Pattern Recognit Lett. 2020;130:38-45. doi: 10.1016/j.patrec.2018.10.003.
9. He X, Chen Y, Ghamisi P. Heterogeneous Transfer Learning for Hyperspectral Image Classification Based on Convolutional Neural Network. IEEE Trans Geosci Remote Sens. 2020;58:3246-63. doi: 10.1109/TGRS.2019.2951445.
10. Hinton GE, Osindero S, Teh Y-W. A Fast Learning Algorithm for Deep Belief Nets. Neural Comput. 2006;18:1527-54. doi: 10.1162/neco.2006.18.7.1527.
11. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, *et al*. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv. 2017. doi: 10.48550/ARXIV.1704.04861.
12. LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. In: The Handbook of Brain Theory and Neural Networks. Cambridge, MA: MIT Press; 1998. p. 255-8.
13. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998;86:2278-324. doi: 10.1109/5.726791.
14. Li H, Wang C, Cui Y, Hodgson M. Mapping salt marsh along coastal South Carolina using U-Net. ISPRS J Photogramm Remote Sens. 2021;179:121-32. doi: 10.1016/j.isprsjprs.2021.07.011.
15. Li J, Xi B, Li Y, Du Q, Wang K. Hyperspectral Classification Based on Texture Feature Enhancement and Deep Belief Networks. Remote Sens. 2018;10:396. doi: 10.3390/rs10030396.
16. Li S, Song W, Fang L, Chen Y, Ghamisi P, Benediktsson JA. Deep Learning for Hyperspectral Image Classification: An Overview. IEEE Trans Geosci Remote Sens. 2019;57:6690-709. doi: 10.1109/TGRS.2019.2907932.
17. Li Y, Hou Z, Zhang L, Qu Y, Zhou G, Lin J, *et al*. Long-term spatio-temporal changes of wetlands in Tibetan Plateau and their response to climate change. Int J Appl Earth Obs Geoinf. 2023;121:103351. doi: 10.1016/j.jag.2023.103351.
18. Li Y, Yu X, Zhang J, Zhang S, Wang X, Kong D, *et al*. Improved Classification of Coastal Wetlands in Yellow River Delta of China Using ResNet Combined with Feature-Preferred Bands Based on Attention Mechanism. Remote Sens. 2024;16:1860. doi: 10.3390/rs16111860.
19. Lin X, Cheng Y, Chen G, Chen W, Chen R, Gao D, *et al*. Semantic Segmentation of China's Coastal Wetlands Based on Sentinel-2 and Segformer. Remote Sens. 2023;15:3714. doi: 10.3390/rs15153714.
20. Liu P, Zhang H, Eom KB. Active Deep Learning for Classification of Hyperspectral Images. IEEE J Sel Top Appl Earth Obs Remote Sens. 2017;10:712-24. doi: 10.1109/JSTARS.2016.2598859.
21. Luo J, He Z, Lin H, Wu H. Biscale Convolutional Self-Attention Network for Hyperspectral Coastal Wetlands Classification. IEEE Geosci Remote Sens Lett. 2024;21:1-5. doi: 10.1109/LGRS.2024.3351551.
22. Ma T, Li X, Bai J, Ding S, Zhou F, Cui B. Four decades' dynamics of coastal blue carbon storage driven by land use/land cover transformation under natural and anthropogenic processes in the Yellow River Delta, China. Sci Total Environ. 2019;655:741-50. doi: 10.1016/j.scitotenv.2018.11.287.
23. Mcleod E, Chmura GL, Bouillon S, Salm R, Björk M, Duarte CM, *et al*. A blueprint for blue carbon: toward an improved understanding of the role of vegetated coastal habitats in sequestering CO2. Front Ecol Environ. 2011;9:552-60. doi: 10.1890/110004.
24. Mohammadimanesh F, Salehi B, Mahdianpari M, Motagh M. A New Hierarchical Object-Based Classification Algorithm for Wetland Mapping in Newfoundland, Canada. In: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium; 2018. p. 9233-6. doi: 10.1109/IGARSS.2018.8517844.
25. Mou X, Liu X, Yan B. Classification System of Coastal Wetlands in China. Wetl Sci. 2015;13(1):19-26. doi: 10.13248/j.cnki.wetlandsci.2015.01.004.
26. Rasel SMM, Chang H-C, Diti IJ, Glasby T. Support Vector Machine (SVM) Classifier with Small Training Samples for Mapping Saltmash Wetland at Species Level. In: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium; 2019. p. 2674-7. doi: 10.1109/IGARSS.2019.8897942.
27. Roy SK, Krishna G, Dubey SR, Chaudhuri BB. HybridSN: Exploring 3-D–2-D CNN Feature Hierarchy

for Hyperspectral Image Classification. IEEE Geosci Remote Sens Lett. 2020;17:277-81. doi: 10.1109/LGRS.2019.2918719.
28. Rumelhart DE, McClelland JL. Learning Internal Representations by Error Propagation. In: Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations. Cambridge, MA: MIT Press; 1987. p. 318-62.
29. Salas EAL, Kumaran SS, Bennett R, Willis LP, Mitchell K. Machine Learning-Based Classification of Small-Sized Wetlands Using Sentinel-2 Images. AIMS Geosci. 2024;10:62-79. doi: 10.3934/geosci.2024005.
30. Santara A, Mani K, Hatwar P, Singh A, Garg A, Padia K, *et al*. BASS Net: Band-Adaptive Spectral-Spatial Feature Learning Neural Network for Hyperspectral Image Classification. IEEE Trans Geosci Remote Sens. 2017;55:5293-301. doi: 10.1109/TGRS.2017.2705073.
31. Wang F, Sanders CJ, Santos IR, Tang J, Schuerch M, Kirwan ML, *et al*. Global blue carbon accumulation in tidal wetlands increases with climate change. Natl Sci Rev. 2021;8:nwaa296. doi: 10.1093/nsr/nwaa296.
32. Yang X, Ye Y, Li X, Lau RYK, Zhang X, Huang X. Hyperspectral Image Classification With Deep Learning Models. IEEE Trans Geosci Remote Sens. 2018;56:5408-23. doi: 10.1109/TGRS.2018.2815613.
33. Yu J, Wang Y, Li Y, Dong H, Zhou D, Han G, *et al*. Soil organic carbon storage changes in coastal wetlands of the modern Yellow River Delta from 2000 to 2009. Biogeosciences. 2012;9:2325-31. doi: 10.5194/bgd-9-1759-2012.
34. Zhao H, Lin Y, Delang CO, Ma Y, Zhou J, He H. Contribution of soil erosion to the evolution of the plateau-plain-delta system in the Yellow River basin over the past 10,000 years. Palaeogeogr Palaeoclimatol Palaeoecol. 2022;601:111133. doi: 10.1016/j.palaeo.2022.111133.
35. Zun-You K, Ru A, Xiang-Juan L. ANN Based High Spatial Resolution Remote Sensing Wetland Classification. In: 2015 14th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES); 2015. p. 180-3. doi: 10.1109/DCABES.2015.52.

How to Cite This Article
Wu W, Kong D, Tauqir M, Zhang J. Accurate classification of coastal wetland in Yellow River Delta based on ResNet combined with coordinate attention. Int J Artif Intell Eng Transform. 2025;6(2):64–75. doi:10.54660/IJAIET.2025.6.2.64-75.

Creative Commons (CC) License
This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.