Check for updates

Research on coastline extraction and dynamic change from remote sensing images based on deep learning

Qingzhe Lv[1], Qi Wang[2], Xiaoli Song[3], Binfu Ge[1], Hao Guan[1], Tongtong Lu[1] and Zui Tao[2]*

[1]School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu, China, [2]Aerospace Information Research Institute, Chinese Academy of Sciences (CAS), Beijing, China, [3]Department of Economic Management, Beijing City University, Beijing, China

Accurate coastline extraction is crucial for the scientific management and protection of coastal zones. Due to the diversity of ground object details and the complexity of terrain in remote sensing images, the segmentation of sea and land faces challenges such as unclear segmentation boundaries and discontinuous coastline contours. To address these issues, this study improve the accuracy and efficiency of coastline extraction by improving the DeepLabv3+ model. Specifically, this study constructs a sea-land segmentation network, DeepSA-Net, based on strip pooling and coordinate attention mechanisms. By introducing dynamic feature connections and strip pooling, the connection between different branches is enhanced, capturing a broader context. The introduction of coordinate attention allows the model to integrate coordinate information during feature extraction, thereby allowing the model to capture longer-distance spatial dependencies. Experimental results has shown that the model can achieves a land-sea segmentation mean intersection over union (mIoU) ration and Recall of over 99% on all datasets. Visual assessment results show more complete edge details of sea-land segmentation, confirming the model's effectiveness in complex coastal environments. Finally, using remote sensing data from a coastal area in China as an application instance, coastline extraction and dynamic change analysis were implemented, providing new methods for the scientific management and protection of coastal zones.

KEYWORDS

sea-land segmentation, remote sensing images, deep learning, strip pooling, coordinate attention

1 Introduction

    With the development of remote sensing technology, the means of obtaining remote sensing data have become more diversified, making coastline extraction from remote sensing images an important means of monitoring coastline changes. Accurate coastline extraction is the foundation for monitoring changes in coastal zones, providing direct data support for understanding the evolutionary trends of coastal zones (Yasir et al., 2020). The accuracy of coastline extraction directly affects coastal zone planning, engineering construction, disaster prevention and mitigation, etc. (Zheng et al., 2023). By regularly acquiring remote sensing image data and accurately extracting and analyzing the dynamic changes of coastlines, it is helpful to identify phenomena such as coastline erosion and

siltation, providing scientific basis for relevant decisions. In addition, the dynamic changes of coastlines are of great significance for studying marine ecological environment and climate change (Ranasinghe et al., 2012). By extracting coastlines and analyzing changes in historical images, it is possible to discover long-term migration trends of coastlines, providing data support for land use, ecological protection, disaster warning, etc. in coastal areas.

However, the process of extracting coastlines still faces multiple challenges, especially in high-resolution remote sensing images. The diversity of ground features and the complexity of terrain often lead to errors in using traditional methods for coastline extraction, such as misidentifying the boundary of closed waters near the coast, such as aquaculture ponds, as sea areas (Tsokos et al., 2018). In addition, coastlines themselves include various types such as artificial, sandy, and silty, and are easily disturbed by nearby ground features. These factors often cause gaps in the extracted coastline, affecting the accuracy of extraction. Traditional methods for coastline extraction can be generally classified into three categories: threshold segmentation-based, classification-based, and edge detection-based methods. In practical applications, due to the challenges of selecting thresholds in complex images, these methods have certain limitations in accuracy, especially in complex coastal environments. At the same time, due to the large amount, diverse types, and rich bands of currently available remote sensing image data, traditional coastline extraction algorithms are difficult to process and analyze massive remote sensing data quickly and accurately, hindering the development of practical applications.

In recent years, deep learning algorithms have been widely used in various computer vision tasks due to their powerful feature extraction capabilities, and have also provided new methods for extracting coastlines from remote sensing images (Bengoufa et al., 2021). The emergence of fully convolutional networks and skip connection models based on Encoder-Decoder structures has further improved the accuracy and efficiency of deep learning in remote sensing image analysis tasks (Han et al., 2018). For instances, Zhou et al. (2018) proposed the D-LinkNet model based on pre-trained encoders and dilated convolutions to extract ground objects from high-resolution satellite images, improving the accuracy of feature extraction and segmentation. Zhu et al. (2021) proposed a dual-branch encoding and iterative attention decoding network specifically for semantic segmentation of remote sensing images, using skip connections to refine the segmentation results. Due to deep learning methods can make full use of multi-scale information in remote sensing images through feature extraction and learning in multiple layers of networks, they have significant advantages to make land-sea boundary recognition more accurate, especially in complex terrains and environments.

Remote sensing images with different spectral bands carry unique spectral information, providing rich features for coastline extraction. Deep learning algorithms can effectively utilize different spectral band data from remote sensing images by constructing convolutional neural networks and other advanced models. Through the input of multi-band spectral information, deep learning models can extract advanced features to analyze the spectral differences between water bodies and other ground objects. These spectral features combined with multi-level feature extraction methods further improve the segmentation accuracy in complex terrain and diverse environments. By using multi-band spectral information, deep learning models can better identify the characteristics of land and sea under different coastal conditions such as sandy, artificial, and silty conditions, not only improving the accuracy of coastline extraction, but also enhancing the generalization ability of the model.

Overall, deep learning technology provides a new approach for coastline extraction. The main contributions of this study are as follows:

(1) Constructed a novel DeepSA-Net remote sensing image segmentation network. By introducing strip pooling and coordinate attention in the model construction, the dynamic feature connection capability during feature extraction is enhanced, enabling the capture of longer-distance spatial dependencies, thereby segmenting more complete and accurate sea-land edge details. This effectively addresses the challenges of complex coastal environments. The proposed DeepSA-Net model improves the accuracy and efficiency of sea-land segmentation in remote sensing images, better serving coastline extraction work.

(2) Verified the application of DeepSA-Net in actual coastline change scenarios. Experimental results show that the model's sea-land segmentation mIoU and Recall rate on the adopted dataset both exceed 99%, and the model's effectiveness in practical applications was validated using remote sensing data from a coastal area in China. This facilitates the analysis of dynamic coastline changes, providing a new method for the scientific management and protection of coastal zones.

2 Related works

Coastline detection essentially combines semantic segmentation and edge detection tasks for dual marine pixel identification. Specifically, the research on coastline extraction methods has gone through multiple stages, including remote sensing index methods, threshold methods, edge detection methods, and machine-learning based methods. Each type of methods has its own advantages and limitations. Thus, Zhou et al. (2023) systematically summarized the development history of coastline extraction methods based on remote sensing dataset. In this section, we discuss traditional coastline extraction methods and deep learning-based coastline extraction methods, starting with coastline extraction.

2.1 Traditional coastline extraction techniques

Traditional methods for coastline detection mainly rely on expert knowledge to analyze samples' spectral, color, geometric, and texture features to distinguish targets from the background. Current traditional coastline extraction methods can be summarized into three categories: (1) threshold-based methods, (2) classification-based methods, and (3) edge detection-based methods.

Threshold-based methods classify target and background objects based on pixel points through histogram analysis. The

normalized difference water index (NDWI) is an important metric of the normalized ratio index of near-infrared and short-wave infrared bands, used in threshold segmentation methods (McFeeters, 1996). However, the extraction process of NDWI is prone to mixing other information, leading to suboptimal extraction accuracy. To resolve this problem, Xu (2005) proposed the modified normalized difference water index (MNDWI) to optimize the extraction of water body boundaries, but challenges remain in distinguishing water bodies from vegetation. To this, Sharma (Sharma et al., 2015) used the superfine water index (SWI) to improve the detail of water body identification. However, selecting thresholds in complex images remains challenging, leading to classification errors.

Classification-based methods are divided into pixel-based classification and object-based classification. To the former, the K-means algorithm (Shang et al., 2022) is widely used for remote sensing image sea-land segmentation tasks. However, selecting the K value in practical applications is difficult, so Liu et al. (2011) introduced the iterative self-organizing data (ISODATA) analysis technique to improve this algorithm. Toure et al. (2018) adopted the fusion of over segmentation (FOOS) technique to identify categories of waves and beaches. To the latter, object analysis does not study single images independently but deals with collections of image pixels, emphasizing the spectral and spatial characteristics of images. Zhang et al. (2013) analyzed coastline images of different resolutions through the new feature of the object merging index (OMI).

Edge detection-based methods locate edge points by identifying color and grayscale changes in images and generate sea-land segmentation contours. Sobel (Wang et al., 2005) and Laplacian operators (Coleman et al., 2010) have been widely used in such tasks, but the Laplacian has limitations in directionality. Asaka et al. (2013) improved this issue with the Laplacian of Gaussian (LoG) operator. Additionally, the Canny operator has been used in coastline detection, for example, Li et al. (2013) extracted the coastline of Bohai Bay using the Canny operator.

In summary, traditional methods for coastline detection are often simple to operate, easy to understand and implement, but may overly rely on expert experience for threshold selection, thus lacking generalizability. These methods are also easily affected by environmental factors such as clouds and vegetation, leading to misclassification. In practical applications, it is usually necessary to combine multiple methods to improve detection accuracy and adaptability.

2.2 Deep learning-based coastline extraction techniques

The powerful end-to-end processing capability of deep learning technology and its significant advantage in directly extracting high-level semantic features have played a huge role in fields such as computer vision and natural language processing. For instances, Yu et al. (2017) considering the spectral and spatial information of images, developed a framework combining convolutional neural networks (CNN) and logistic regression, effectively extracting coastlines from Landsat-7 satellite images. To reduce the loss of image features during the pooling process, Chen et al. (2018a) used a simple linear iterative clustering algorithm to divide images into

superpixels, then combined a new CNN model to extract urban water bodies and detect building shadows.

Although CNNs excel in extracting local features of images, their classification performance is limited by the size of the receptive field. To overcome this limitation, Shelhamer et al. (2017) proposed the fully convolutional networks (FCN), which replaces fully connected layers with upsampling techniques, allowing for processing images of any size and demonstrating superior performance. However, FCN may lead to unclear segmentation details and issues with spatial consistency and boundary blurriness during convolution and pooling operations. To address these issues, scholars introduced Encoder-Decoder structured models with skip connections, such as DeepLabv3+ (Chen L. et al., 2018) and UNet (Ronneberger et al., 2015). These models preserve image contour information by retaining pooling coordinates, thus mitigating information loss caused by pooling operations in FCN. Particularly, the UNet model achieves precise localization by adding a contracting path in the Encoder-Decoder module, which performs well in fields such as biomedical and coastline detection.

With technological advancements, methods combining standard convolutional layers with dense connections, dilated convolutions, residual modules, and multi-scale feature pyramids have been widely adopted. For examples, Cao et al. (2020) developed the DenseUNet model based on UNet, extending the DenseNet idea to UNet for image segmentation, and proposed a weighted loss to accurately obtain segmentation results and overcome gradient vanishing issues. Wang et al. (2022) designed a novel upsampling convolution-deconvolution module, thereby improving the performance of remote sensing image classification. Furthermore, Cheng et al. (2017) optimized multi-task models' segmentation and edge detection through edge-aware regularization, enhancing the classification accuracy for oceans, land, and ships. Li et al. (2020) proposed an attention mechanism-based convolutional neural network with multiaugmented schemes, which effectively reduces the information redundancy introduced during feature extraction by forcing the model to capture features of specific classes, and improves the recognition performance of remote sensing images.

These advancements demonstrate the potential and effectiveness of fully convolutional symmetric semantic segmentation models, especially networks such as DeepLab and U-Net, in applications like coastline detection. By drawing on these advanced research ideas and technologies, this study focuses on deep learning-based coastline detection from the perspectives of pooling methods and attention mechanisms, based on the DeepLabv3+ model.

3 Methodology

3.1 Network architecture

DeepLabv3+ is a neural network model for image semantic segmentation tasks, mainly addressing the blurriness issue in processing sharp object edges in earlier versions by introducing an Encoder-Decoder structure. In the Encoder part, DeepLabv3+ uses the Xception network as the backbone network, optimizing computational efficiency and model performance with depthwise separable convolutions. Additionally, the model integrates the

FIGURE 1
Network structure diagram of DeepSA-Net.
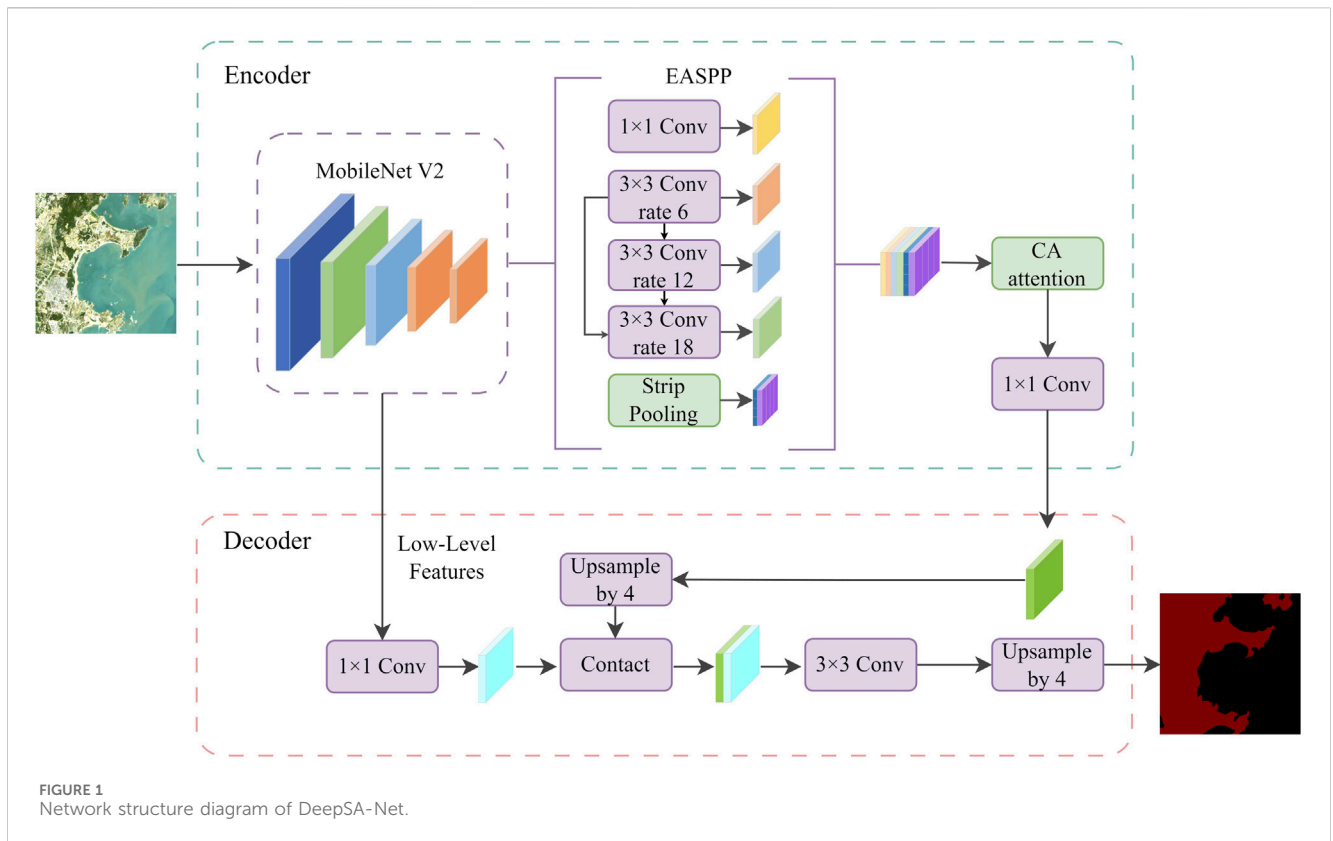
atrous spatial pyramid pooling (ASPP) module, consisting of several convolutional layers with different dilation rates and a 1×1 convolutional layer, to capture multi-scale features, effectively extracting high-level semantic information and adapting to different task requirements. In the Decoder part, the model fuses deep and shallow features extracted by the Encoder using channel concatenation technology and uses upsampling to restore the feature image to the original size of the input image, achieving precise pixel-level semantic segmentation. This structure not only enhances the network's ability to capture image details but also improves recognition of sharp edges. Due to these advantages, DeepLabv3+ has shown strong adaptability and efficiency in fields such as medical image segmentation.

However, in the task of sea-land segmentation of remote sensing images, the fixed branch structure of the traditional ASPP module limits its ability to handle complex scenes, and it can only capture features within a rectangular space, lacking in directional feature extraction. This study introduces a dynamic feature connection structure and integrates the strip pooling module, proposing an EASPP (Enhanced ASPP) module. On the one hand, the inclusion of this module allows the outputs of different convolutional branches to be reused in subsequent branches, enhancing the connection between different branches. On the other hand, by performing adaptive average pooling on the width and height of the image, it more effectively captures global features in horizontal and vertical directions. This method better captures and integrates features from different spatial scales, providing a richer and more detailed feature representation for complex sea-land interfaces in remote sensing images, such as winding coastlines and water bodies of various sizes.

Furthermore, the original DeepLabv3+ model lacks a dedicated attention mechanism. Although ASPP can handle features from different scales, its processing is still limited to relatively independent dilated convolution branches, with insufficient interaction between channels. This structure is not sufficient to capture long-distance spatial dependencies, affecting the overall segmentation's coherence and accuracy. This study introduces the coordinate attention module, encoding the input feature map in spatial directions, allowing the model to capture longer spatial dependencies. By integrating coordinate information while focusing on the relationship between feature map channels, the model can more finely adjust the contribution of each channel, thereby enhancing the response to specific features and improving segmentation accuracy and robustness.

Based on these improvements, this study proposes a DeepSA-Net model based on DeepLabv3+, as shown in Figure 1. Below, we will provide a detailed introduction to the core modules that construct the DeepSA-Net network.

3.2 EASPP module

In the field of image segmentation, ASPP has been proven to be an effective method for capturing image features at different spatial scales. ASPP enhances the model's perception of features at different scales while maintaining computational efficiency through a series of parallel dilated convolutional layers with different dilation rates. However, the traditional ASPP structure mainly uses a fixed branch structure to process input features and fuses the outputs of these branches through a unified convolutional layer, which still has room
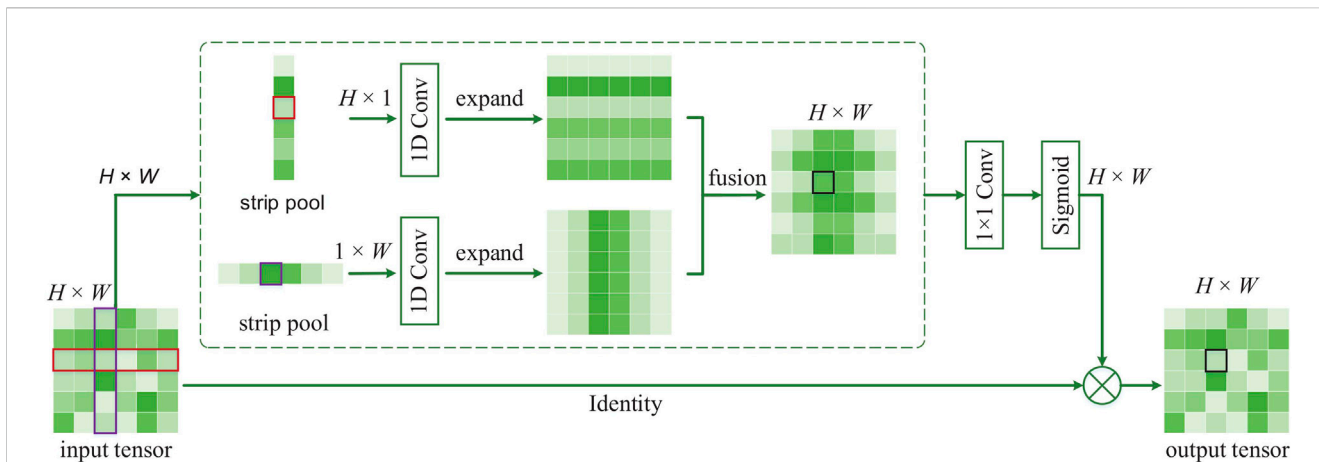
FIGURE 2
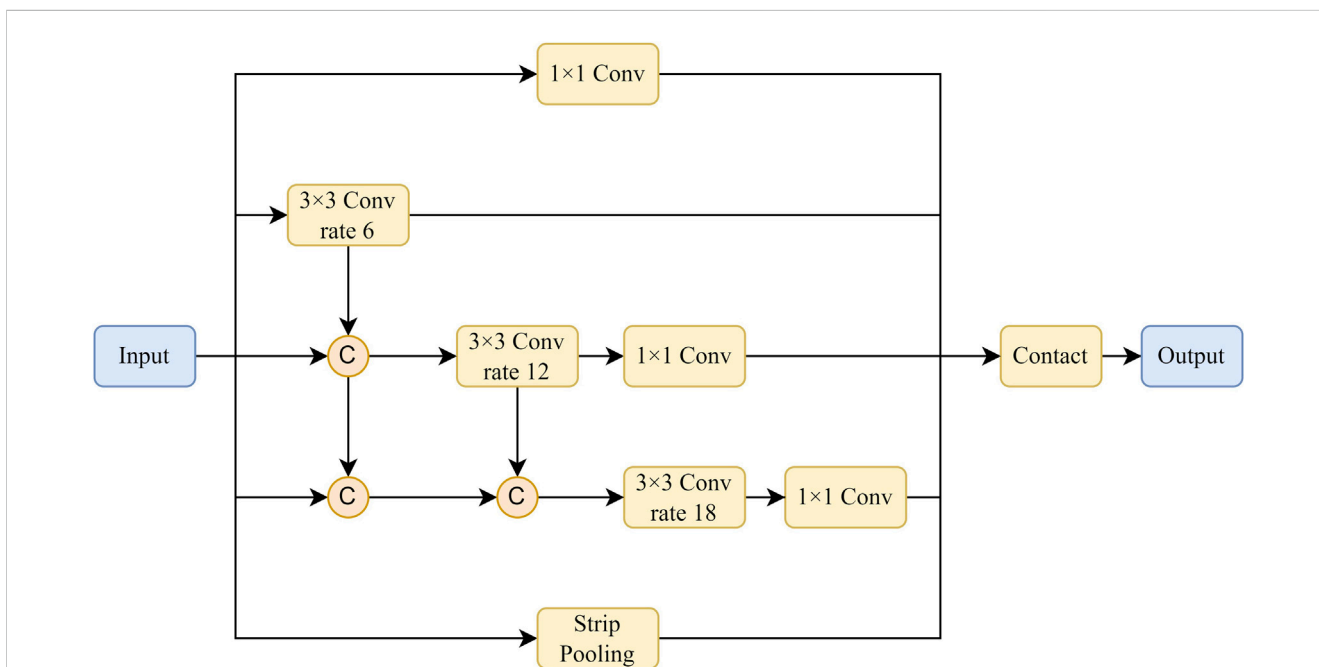Strip pooling module structure diagram.

FIGURE 3
EASPP module structure diagram.

for improvement in handling complex scenes. In addition, the ASPP method only detects input feature maps through square windows, limiting its flexibility in capturing anisotropic features widely present in remote sensing images.

To address these limitations, this study proposes an enhanced ASPP module, named EASPP. The new module introduces a dynamic feature connection structure, allowing the outputs of different convolutional branches to be reused in subsequent branches, thereby enhancing the connection between different branches and enriching the context information the model can utilize. Specifically, some branches in EASPP combine the outputs of previous branches with the original input before performing dilated convolution. For example, the third branch combines the output of the second branch with the original input, and the fourth branch

connects the outputs of the third and second branches with the original input. This method allows each branch to operate not just independently on its input but in a richer feature environment, making feature extraction more refined and in-depth.

Furthermore, EASPP integrates the strip pooling module, designed to address the issue of insufficient feature capture in image segmentation. Thus, the EASPP module, through unique strip pooling operations, performs adaptive average pooling on the width and height dimensions of the image, capturing global contextual information in different directions. This process helps the model better understand the overall structure of the image and enhances its grasp of spatial details. After effectively extracting global features in the horizontal and vertical directions of remote sensing images, these features are fused and added to the original

input features, enhancing feature representativeness while preserving original feature information, allowing the model to perform segmentation tasks more accurately in complex environments.

The structure diagram of the strip pooling module is shown in Figure 2. As shown in Figure 2, compared with traditional pooling methods, strip pooling pools and obtains one-dimensional feature vectors through horizontal and vertical paths, and then obtains the outputs result through resolution expansion and feature map fusion. Specifically, the module first receives a $C×H×W$ feature map, then processes it through horizontal and vertical strip pooling, outputting two feature maps of sizes $H × 1$ and $1×W$. By averaging the elements within the pooling kernel, the pooled outputs are obtained. Next, through convolution operations, these two pooled feature maps are expanded along their width and height, respectively, to the same size, and the two feature maps are summed pixel-wise to obtain an $H×W$ fused feature. Finally, a $1 × 1$ convolution operation and Sigmoid processing yield the final output. This design allows the strip pooling module to capture extensive contextual information without losing sensitivity to local features.

In the task of sea-land segmentation of remote sensing images, effectively processing and distinguishing different geographical spatial features is crucial. Since oceans and large water bodies typically exhibit significant area continuity, depending on the specific geographical direction and coastline orientation, this continuity may be more pronounced in either the vertical or horizontal direction. Strip pooling model, by performing horizontal and vertical adaptive average pooling, can effectively extract global features in these directions. Such processing not only helps to more accurately delineate the boundaries between water and land but also allows the model to handle different coastline orientations more flexibly and accurately. Additionally, this strategy improves the model's segmentation performance for terrain features such as rivers and islands, as these features' representation in images is often closely related to their geographical direction. Through this method, strip pooling model enhances the spatial understanding capability of remote sensing image segmentation models, especially in complex natural and artificial interface-interlaced coastal areas.

Overall, EASPP, by integrating dynamic feature connections and additional global information provided by strip pooling and integrating them through $1 × 1$ convolution, results in a more comprehensive feature representation. This strategy not only retains the detail information obtained from different dilation rate convolutional branches, but also strengthens the integration of these details with additional global features, making the final feature representation more comprehensive and effective. The structure diagram of the EASPP module is shown in Figure 3.

The task of sea-land segmentation of remote sensing images often involves processing highly heterogeneous landscapes, including complex coastlines and diverse boundaries between water bodies and land. Through a multi-scale atrous convolution strategy, the EASPP model can capture detailed contextual information at different spatial scales. EASPP model, with its dynamic feature fusion capability and integration of global information, greatly enhances the model's accuracy and robustness in handling these complex scenes. Particularly, the introduction of strip pooling increases the model's ability to capture features in broad and elongated areas along the sea-land boundary, thus providing more precise and coherent sea-land segmentation results.

3.3 Coordinate attention module

Currently, many lightweight neural network architectures have adopted the squeeze-and-excitation (SE) module to implement attention mechanisms, focusing on the interaction between channels but paying less attention to spatial location information. To overcome this limitation, convolutional block attention module (CBAM) adds consideration of spatial location on top of channel attention, using convolution operations after channel reduction to extract location attention information. However, convolution operations typically capture only local feature relationships, lacking consideration for long-distance relationships. For this reason, the coordinate attention mechanism was proposed. Coordinate attention can be seen as an optimized computational unit designed to enhance the network's feature expression capability. It not only recalibrates features at the channel level but also enhances the network's feature expression by embedding spatial location information.

This mechanism processes input feature groups $[x_1, x_2, \cdots, x_C] \in \mathbb{R}^{C×H×W}$, to produce enhanced features $[y_1, y_2, \cdots, y_C] \in \mathbb{R}^{C×H×W}$ with the same dimensions as the input. It mainly consists of two parts: the coordinate information embedding module and the coordinate attention generation module, which work together to make the network more efficient and precise in global perception and local detail capture. The module structure is shown in Figure 4.

(1) Coordinate Information Embedding.

First, pooling kernels of size of $(1, W)$ and $(H, 1)$ are used to encode each channel of the input feature map along the vertical and horizontal directions, respectively. Suppose $x_i(h, w)$ represents the element at position (h, w) in the ith channel of the feature map. During vertical encoding, the size of each
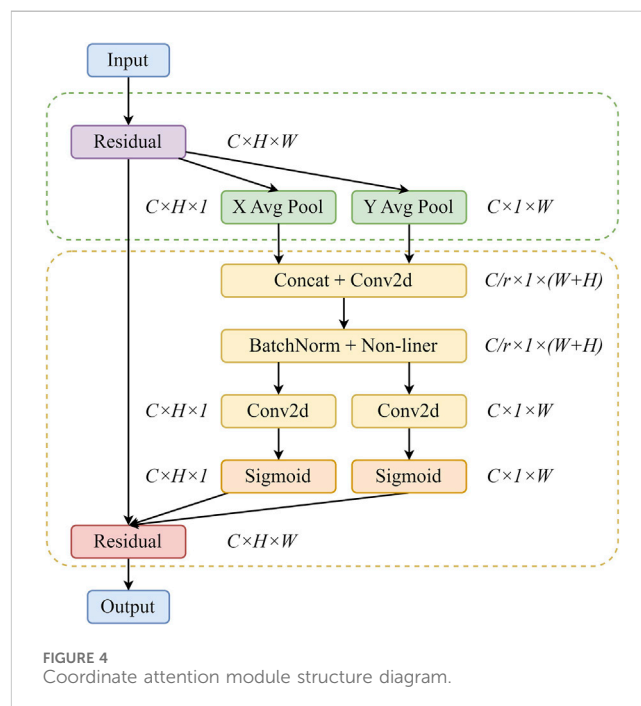

FIGURE 4
Coordinate attention module structure diagram.

channel feature map becomes $H \times 1$, and the output at height h in the ith channel is shown in Equation 1:

$$z_i^h = \frac{1}{W} \sum_{n=0}^{W-1} x_i(h, n) \tag{1}$$

During horizontal encoding, the size of each channel feature map becomes $1 \times W$, and the output at width w in the ith channel is shown in Equation 2.

$$z_i^w = \frac{1}{H} \sum_{m=0}^{H-1} x_i(m, w) \tag{2}$$

After the above pooling operations, image features are aggregated in two spatial directions, vertical and horizontal, producing a pair of feature maps. Each feature map contains long-distance dependencies along one spatial direction and preserves precise location information along the other spatial direction. Embedding coordinate information into feature maps helps the network more accurately locate targets of interest.

(2) Coordinate Attention Generation.

The coordinate attention generation process involves concatenating the feature maps after coordinate information embedding and using a 1×1 convolution to adjust the number of channels in the feature map, as shown in Equation 3.

$$f = \delta\left(F_1\left(\left[z^h, z^w\right]\right)\right) \tag{3}$$

where, F_1 represents a 1×1 convolution operation, δ is a nonlinear activation function, z^h and z^w represents the feature vector after coordinate information embedding. The generated feature tensor is $f \in \mathbf{R}^{C/r \times 1 \times (H+W)}$, where r is the channel downsampling rate. The feature tensor f is then split into two independent tensors $f^h \in \mathbf{R}^{C/r \times H \times 1}$ and $f^w \in \mathbf{R}^{C/r \times 1 \times W}$, along the spatial dimension. Then, two independent 1×1 convolutions F_h and F_w are used to restore the channel count of feature vectors F_h and F_w to C, as shown in Equations 4, 5.

$$g^h = \sigma\left(F_h\left(f^h\right)\right) \tag{4}$$

$$g^w = \sigma\left(F_w\left(f^w\right)\right) \tag{5}$$

where σ is the *sigmoid* activation function, thus obtaining the attention weights $g^h \in \mathbf{R}^{C \times H \times 1}$ and $g^w \in \mathbf{R}^{C \times 1 \times W}$. The obtained attention weights are used to update the input feature map X to produce the output feature map $Y \in \mathbf{R}^{C \times H \times W}$, as shown in Equation 6.

$$y_c(i, j) = x_c(i, j) \times g_c^h(i, 1) \times g_c^w(1, j) \tag{6}$$

where, $y_c(i, j)$ and $x_c(i, j)$ are the values of the cth channel of the output feature map and the original feature map at position of (i, j), $g_c^h(i, 1)$ and $g_c^w(1, j)$ are the values of the cth channel of the attention weights generated along the vertical and horizontal directions at position of (i, j), respectively. Through Formula 6, it combines the input feature map with the attention weights to generate an output feature map that highlights important features, thereby improving the model's performance in spatial and long-distance dependencies.

In summary, coordinate attention can supplement the shortcomings of DeepLabv3+ in understanding spatial details and long-distance dependencies, and by utilizing coordinate information to locate areas of interest, it allows the model to more effectively capture long-distance spatial dependencies, fully utilizing information between channels to enrich feature expression.

4 Experiment and discussion

4.1 Datasets and data processing

This study primarily utilized the SLS (sea-land segmentation) dataset collected by Yang et al. (2020) in 2020, consisting of 29 Landsat-8 OLI (operational land imager) images covering coastal areas of China. For neural network input convenience, all remote sensing images were segmented into 3361 image blocks of 512×512 pixels, with 2689 blocks used as training samples and 672 as test samples. Each small image block was manually labeled
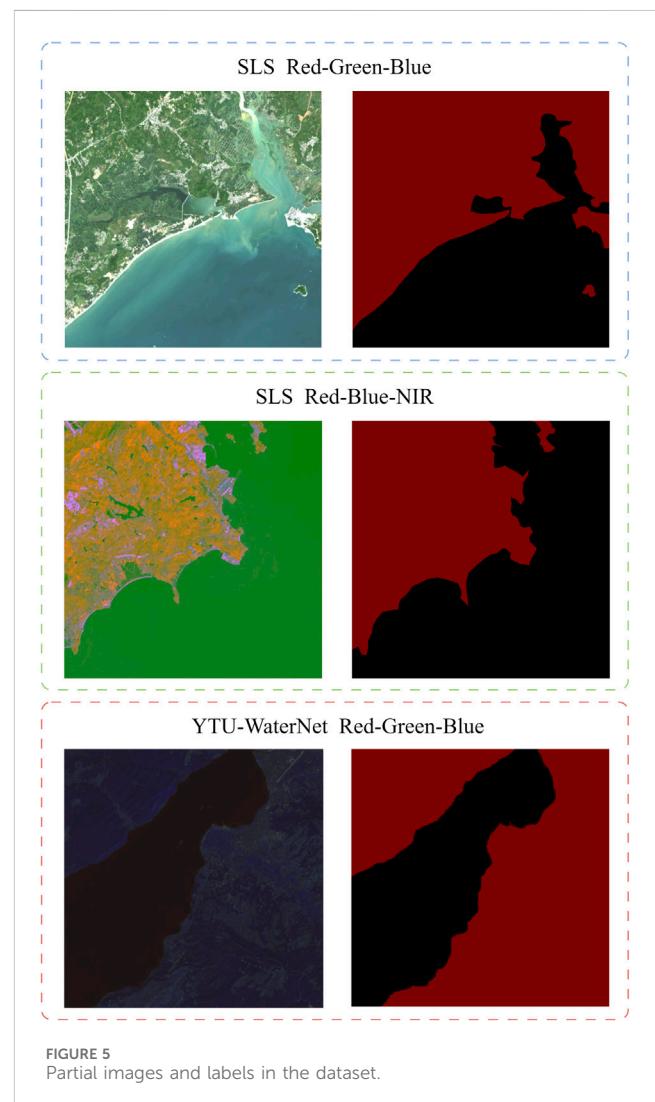

FIGURE 5
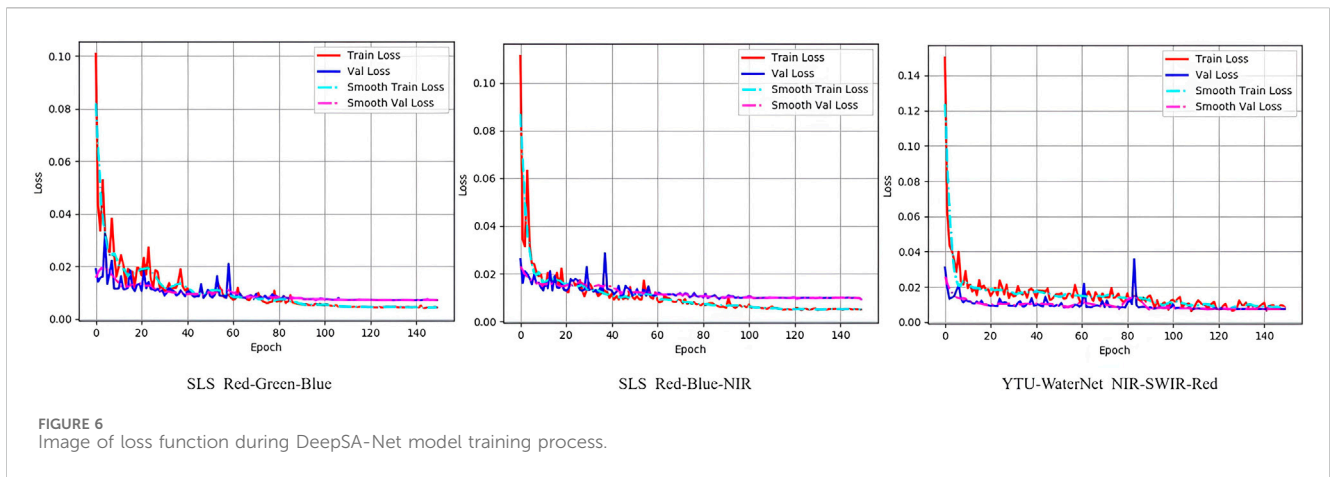Partial images and labels in the dataset.

FIGURE 6
Image of loss function during DeepSA-Net model training process.

TABLE 1 Sea land segmentation indicators of five models on three different band datasets.

Dataset	Band of wave	Model	mIoU	Recall	Precision	Accuracy
SLS	Red-Green-Blue	U-Net	98.25	98.45	99.51	99.14
		SCUNet	98.30	98.65	99.37	99.16
		DenseUNet	98.78	99.13	99.45	99.40
		DeepLabv3+	99.02	**99.48**	99.50	99.51
		DeepSA-Net	**99.31**	99.46	**99.74**	**99.66**
	Red-Blue-NIR	U-Net	98.15	98.85	99.04	99.08
		SCUNet	98.25	98.51	99.29	99.21
		DenseUNet	98.53	98.86	99.46	99.27
		DeepLabv3+	98.81	99.15	99.50	99.41
		DeepSA-Net	**99.07**	**99.34**	**99.60**	**99.54**
YTU-WaterNet	NIR-SWIR-Red	U-Net	98.95	99.24	99.63	99.48
		SCUNet	99.06	99.26	**99.74**	99.53
		DenseUNet	98.43	98.72	99.58	99.21
		DeepLabv3+	99.14	**99.58**	99.51	99.57
		DeepSA-Net	**99.37**	99.56	99.73	**99.69**

with two types of labels: land and sea, using Labelme. The study created two sub-datasets using RGB and NIR-SWIR-Red (near infrared - short wave infrared - red) spectral band combinations for training the models. All samples were in PNG format without spatial information. To further assess the model's performance, the study also incorporated the YTU-WaterNet (Yildiz Technical University) dataset proposed by Erdem et al. (2021), consisting of 63 Landsat-8 OLI images covering parts of Europe, South America, North America, and Africa. Only the red, blue, and near-infrared (NIR) bands were used as samples for this dataset. All remote sensing images were segmented into 1008 image blocks of 512 × 512 pixels, with 806 blocks used as training samples and 202 as test samples. Part of the dataset images and labels are shown in Figure 5, where red represents land areas and black represents sea areas.

4.2 Experimental parameters and evaluation indices

All experiments in this study were implemented on the PyTorch framework and conducted on a server equipped with a 40 GB NVIDIA A100 GPU. To ensure consistency in training conditions, both the original and improved models were set with the same hyperparameters: the optimizer was Adam, the initial learning rate was set at $lr = 0.0005$, the minimum learning rate was 0.01 of the initial rate, and the learning rate adjustment strategy was cosine decay. The batch size was set to 20, the input image size was set to 512 × 512 pixels with 8X downsampling, and the models were trained for 150 epochs, with an evaluation every 5 epochs. Dice loss was used to handle class imbalance, and multi-threading was used to

TABLE 2 Detailed visualization of sea land segmentation results of five models on three different band datasets.

speed up data reading. The model with the highest training accuracy was selected for testing.

To verify the algorithm's performance in sea-land segmentation on high-resolution remote sensing images, four evaluation indices were used: mIoU (mean intersection over union), Recall, Precision, and Accuracy. Their calculation formulas are given in Equations 7–10.
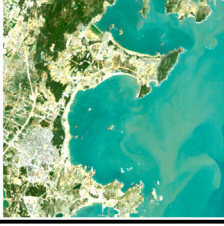
$$mIoU = \frac{1}{2} \times \left(\frac{TP}{TP + FP + FN} + \frac{TN}{TN + FP + FN} \right) \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

TABLE 3 Coastal line extraction results.

Region	Original image	Label	Prediction results	Image with coastline
#1				
#2				
#3				

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (10)$$

where *TP* is the number of sea pixels correctly predicted by the network, *FP* is the number of sea pixels incorrectly predicted, *TN* is the number of land pixels correctly predicted, and *FN* is the number of land pixels incorrectly predicted.

4.3 Experimental results

The study conducted sea-land segmentation experiments on the SLS dataset and YTU-WaterNet dataset using five deep learning models: U-Net, SCUNet, DenseUnet, DeepLabv3+, and DeepSA-Net. Among them, the former dataset includes two band combinations, while the latter dataset only has one band combination. Especially, the loss function of the DeepSA-Net model during the three training processes is shown in Figure 6. As the number of iterations increases, the loss function shows a downward trend before 100 epochs, and the trend gradually stops between 100 and 150 epochs, proving that the model has converged.

The experimental results of land-sea segmentation on three different bands of data sets for the five models are shown in Table 1. In Table 1, the bold font indicates the best metric among all models, and the underline indicates the second best metric among all models.

As shown in Table 1, the training results based on the YTU-WaterNet dataset generally have higher accuracy than those based on the SLS dataset. Through analysis, it can be concluded that the information carried by the NIR and SWIR bands is more conducive to distinguishing between ocean and land. The

training results on all different band datasets show that both DeepLabv3+ and the DeepSA-Net proposed in this study have achieved good results. DeepSA-Net is slightly better than DeepLabv3+ and U-Net models in terms of mIoU, Recall, Precision, and Accuracy. Since the effect of coastline extraction mainly depends on the segmentation effect at the sea-land boundary, these indicators alone may not be sufficient to visually demonstrate the differences between different models. Table 2 selects three regions from each of the above three band datasets and displays their corresponding detailed segmentation results to visually demonstrate the effects of the models. To facilitate observation, the parts with significant differences in the result graph are marked with yellow rectangular boxes.

As shown in Table 2, although the mIoU of each model is above 98%, there are still significant differences in the actual segmentation results of land and sea. DeepLabv3+ and DeepSA-Net proposed in this study perform the best, able to retain more details of the coastline and have good segmentation results for rivers between lands. The results from the third row indicates that the DeepSA-Net and the DeepLabv3+ model outperform other models in terms of prediction accuracy, and especially the DeepSA-Net model closely aligns with the labels (true values) in terms of minor details. From the fifth row of results, it can be seen that only DeepSA-Net and DeepLabv3+ models segment rivers, while other models misclassify rivers as land. From the sixth row of results, it can be seen that for muddy seawater, the DeepSA-Net model still correctly segments islands and land coastlines, showing strong anti-interference ability. From the eighth and ninth rows of results, it can be seen that for sharp details of the coastline in the label, the model proposed in this study retains the

TABLE 4 Analysis of coastal line changes.

Study zone	Time 1 (02/27/2017)	Time 2 (06/29/2023)	Overlaid image with coastline
#1			
#2			
#3			

best effect, while other models have problems with unclear boundary segmentation details and blurred boundaries.

4.4 Coastline extraction

Based on the results of sea-land segmentation with our proposed DeepSA-Net model, the study explored methods for coastline extraction. The coastline was extracted from the label images based on the principle of maximum continuity and regularization methods to fit potential breakpoints in the curve, thus achieving accurate and continuous coastline extraction. The principle of maximum continuity refers to finding the longest approximate curve in the segmentation label results as the coastline extraction result. The regularization method fits the possible breakpoints in the curve based on the coordinates of the previous and subsequent pixels, resulting in an accurate and continuous coastline.

Table 3 shows the coastline extraction results of three regions in the SLS dataset. The original labels and prediction results are processed by the algorithm to generate coastline contour data, and the contour lines are overlaid on the original image. The green curve is the prediction output of the DeepSA-Net network, and the red curve is the original label.

From Table 3, it can be seen that the method proposed in this study extracts a complete and continuous coastline. The two colored curves overlap in most coastal areas, with only a few areas where they are difficult to distinguish, proving the effectiveness of the coastline extraction method.

4.5 Experiment on the analysis of coastline dynamic changes

Based on the above researches, the study conducted coastline extraction and dynamic change analysis in a coastal area near Tangshan City, Hebei Province, China. This area is located near a river estuary where the coastline changes are quite apparent. Remote sensing images from three study zones in this area were downloaded from the ArcGIS Living Atlas of the World platform, taken on 27 February 2017, and 29 June 2023. After processing with the DeepSA-Net model, sea-land segmentation labels were extracted, and the extracted coastline contours were overlaid on the images of 2023 for display, as shown in Table 4. The red line represents the coastline extracted from the images of 2017, and the green line represents the coastline from the images of 2023, allowing for a visual analysis of the changes in the coastline over the years.

As shown in Table 4, the coastlines of the selected study zones have undergone significant changes in 2017 and 2023. Specifically, the coastlines of zones #1 and #3 have extended slightly inward, while the coastline of zone #2 has extended slightly outward, mainly due to the scouring and sediment accumulation of rivers near the estuary. Specifically, the sediment carried by rivers deposits at the estuary and gradually forms new sedimentary landforms, leading to the outward expansion of the coastline. The hydrodynamic characteristics of the water flow near the estuary determine the distribution and accumulation process of sediments, which is a common phenomenon in the change of estuary coastline. On the other hand, tides and coastal currents can erode the coastline, especially during the rising and falling tide processes. The back-and-forth flow of seawater erodes coastal sediments, causing the coastline to shrink towards the land side.

From the experimental results, the proposed method in this study accurately realizes the extraction of coastlines. Compared with traditional coastline extraction algorithms, the deep-learning method proposed in this study has higher accuracy and can effectively distinguish between land and sea in different seasons and different coverage types. It can also better distinguish between man-made water bodies and oceans, thus preventing erroneous segmentation of coastlines. The coastline with more complete details and contours also creates favourable conditions for subsequent large-scale and high-precision analysis of coastline changes. Therefore, this work could provide better technical support for coastline monitoring work.

5 Conclusion and future works

This study, by integrating attention mechanisms and other deep learning theories, constructed a sea-land segmentation network, DeepSA-Net, based on strip pooling and coordinate attention mechanisms. Using remote sensing image datasets from coastal areas in China, the study conducted sea-land segmentation and verified the model's effectiveness across multiple datasets. The experimental results on high-resolution remote sensing image datasets demonstrated that the novel network model achieved excellent segmentation effects, with mIoU and Recall rate exceeding 99% across all datasets. Visual assessment confirmed the model's effectiveness in complex coastal environments. Finally, using remote sensing data from a coastal area in China, the study implemented coastline extraction and dynamic change analysis, providing new methods for the scientific management and protection of coastal zones.

In the future, the researches can further optimize the DeepSA-Net network model, especially to adjust the hyperparameters of Adam, to cope with larger-scale and more complex terrain remote sensing image datasets, thereby improving the model's adaptability under different coastal landforms. At the same time, it is possible to explore the application of multi-source data fusion technology to coastline extraction, combining data from different sensors to further improve the accuracy and robustness of coastline extraction, providing stronger technical support for the scientific management and protection of coastal areas.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

QL: Writing–original draft, Writing–review and editing, Data curation, Methodology. QW: Conceptualization, Writing–original draft. XS: Data curation, Validation, Writing–review and editing. BG: Writing–review and editing, Data curation. HG: Writing–review and editing, Methodology. TL: Writing–original draft, Investigation. ZT: Writing–original draft, Writing–review and editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Asaka, T., Yamamoto, Y., Aoyama, S., Iwashita, K., and Kudou, K. (2013) "Automated method for tracing shorelines in L-band SAR images," in IEEE Conference Proceedings of 2013 Asia-Pacific Conference on Synthetic Aperture Radar (APSAR), Tsukuba, Japan, 23–27 September 2013. doi:10.1109/ultsym.2013.0560

Bengoufa, S., Niculescu, S., Mihoubi, M. K., Belkessa, R., and Abbad, K. (2021). Rocky shoreline extraction using a deep learning model and object-based image analysis. The International Archives of the Photogrammetry. Remote Sens. Spatial Inf. Sci. XLIII-B3-2021, 23–29. doi:10.5194/isprs-archives-xliii-b3-2021-23-2021

Cao, Y., Liu, S., Peng, Y., and Li, J. (2020). DenseUNet: densely connected UNet for electron microscopy image segmentation. *IET Image Process.* 14 (12), 2682–2689. doi:10.1049/iet-ipr.2019.1527

Chen, L., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018b). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Lect. notes Comput. Sci.*, 833–851. doi:10.1007/978-3-030-01234-2_49

Chen, Y., Fan, R., Yang, X., Wang, J., and Latif, A. (2018a). Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning. *Water* 10 (5), 585. doi:10.3390/w10050585

Cheng, D., Meng, G., Xiang, S., and Pan, C. (2017). FusionNet: edge aware deep convolutional networks for semantic segmentation of remote sensing harbor images. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 10 (12), 5769–5783. doi:10.1109/jstars.2017.2747599

Coleman, S. A., Scotney, B. W., and Suganthan, S. (2010). Edge detecting for Range data using Laplacian operators. *IEEE Trans. Image Process.* 19 (11), 2814–2824. doi:10.1109/tip.2010.2050733

Erdem, F., Bayram, B., Bakirman, T., Bayrak, O. C., and Akpinar, B. (2021). An ensemble deep learning based shoreline segmentation approach (WaterNet) from Landsat 8 OLI images. *Adv. Space Res.* 67 (3), 964–974. doi:10.1016/j.asr.2020.10.043

Han, W., Feng, R., Wang, L., and Cheng, Y. (2018). A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification. *ISPRS J. Photogrammetry Remote Sens.* 145, 23–43. doi:10.1016/j.isprsjprs.2017.11.004

Li, F., Feng, R., Han, W., and Wang, L. (2020). An augmentation attention mechanism for High-Spatial-Resolution Remote Sensing Image scene classification. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 13, 3862–3878. doi:10.1109/jstars.2020.3006241

Li, X., Yuan, C., and Li, Y. (2013). Remote sensing monitoring and spatial-temporal variation of Bohai Bay coastal zone. *Remote Sens. Nat. Resour.* 25 (2), 156–163. doi:10.6046/gtzyyg.2013.02.26

Liu, H., Wang, L., Sherman, D. J., Wu, Q., and Su, H. (2011). Algorithmic foundation and Software Tools for extracting shoreline features from remote sensing imagery and LiDAR data. *J. Geogr. Inf. Syst.* 03 (02), 99–119. doi:10.4236/jgis.2011.32007

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* 17 (7), 1425–1432. doi:10.1080/01431169608948714

Ranasinghe, R., Duong, T. M., Uhlenbrook, S., Roelvink, D., and Stive, M. (2012). Climate-change impact assessment for inlet-interrupted coastlines. *Nat. Clim. Change* 3 (1), 83–87. doi:10.1038/nclimate1664

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-NET: convolutional networks for biomedical image segmentation. *Med. Image Comput. Computer-Assisted Intervention (MICCAI)*, 234–241. doi:10.1007/978-3-319-24574-4_28

Shang, R., Liu, M., Jiao, L., Feng, J., Li, Y., and Stolkin, R. (2022). Region-level SAR image segmentation based on edge feature and label Assistance. *IEEE Trans. geoscience remote sens.* 60, 1–16. doi:10.1109/tgrs.2022.3217053

Sharma, R., Tateishi, R., Hara, K., and Nguyen, L. (2015). Developing Superfine water index (SWI) for global water cover Mapping using MODIS data. *Remote Sens.* 7 (10), 13807–13841. doi:10.3390/rs71013807

Shelhamer, E., Long, J., and Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Analysis Mach. Intell.* 39 (4), 640–651. doi:10.1109/tpami.2016.2572683

Toure, S., Diop, O., Kpalma, K., and Maiga, A. S. (2018). Coastline detection using fusion of over segmentation and distance regularization level set evolution. *Int. Archives Photogrammetry, Remote Sens. Spatial Inf. Sci.* XLII-3/W4, 513–518. doi:10.5194/isprs-archives-xlii-3-w4-513-2018

Tsokos, A., Kotsi, E., Petrakis, S., and Vassilakis, E. (2018). Combining series of multi-source high spatial resolution remote sensing datasets for the detection of shoreline displacement rates and the effectiveness of coastal zone protection measures. *J. Coast. Conservation* 22 (2), 431–441. doi:10.1007/s11852-018-0591-3

Wang, L., Xu, H., and Li, S. (2005). Dynamic monitoring of the shoreline changes in xiamen island with its surrounding areas of se china using remote sensing technology. *Remote Sens. Technol. Appl.* 20 (4), 404–410. doi:10.11873/J.ISSN.1004-0323.2005.4.404

Wang, W., Kang, Y., Liu, G., and Wang, X. (2022). SCU-NET: semantic Segmentation Network for learning channel information on remote sensing images. *Comput. Intell. Neurosci.* 2022, 1–11. doi:10.1155/2022/8469415

Xu, H. (2005). A study on information extraction of water body with the modified normalized difference water index (MNDWI). *J. Remote Sens.* 2005 (5), 589–595. doi:10.11834/jrs.20050586

Yang, T., Jiang, S., Hong, Z., Zhang, Y., Han, Y., Zhou, R., et al. (2020). Sea-Land segmentation using deep learning techniques for LandsAT-8 OLI imagery. *Mar. Geod.* 43 (2), 105–133. doi:10.1080/01490419.2020.1713266

Yasir, M., Hui, S., Binghu, H., and Rahman, S. (2020). Coastline extraction and land use change analysis using remote sensing (RS) and Geographic Information System (GIS) technology – a review of the literature. *Rev. Environ. Health* 35 (4), 453–460. doi:10.1515/reveh-2019-0103

Yu, L., Wang, Z., Tian, S., Ye, F., Ding, J., and Kong, J. (2017). Convolutional neural networks for water body extraction from Landsat imagery. *Int. J. Comput. Intell. Appl.* 16 (01), 1750001. doi:10.1142/s1469026817500018

Zhang, T., Yang, X., Hu, S., and Su, F. (2013). Extraction of coastline in aquaculture coast from Multispectral remote sensing images: object-based region Growing integrating edge detection. *Remote Sens.* 5 (9), 4470–4487. doi:10.3390/rs5094470

Zheng, J., Xu, W., Tao, A., Fan, J., Xing, J., and Wang, G. (2023). Synergy between coastal ecology and disaster mitigation in China: Policies, practices, and prospects. *Ocean & Coast. Manag.* 245, 106866. doi:10.1016/j.ocecoaman.2023.106866

Zhou, L., Zhang, C., and Wu, M. (2018) "D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *IEEE/CVF Conference on computer vision and Pattern recognition Workshops*, 192–1924. doi:10.1109/CVPRW.2018.00034

Zhou, X., Wang, J., Zheng, F., Wang, H., and Yang, H. (2023). An Overview of coastline extraction from remote sensing data. *Remote Sens.* 15 (19), 4865. doi:10.3390/rs15194865

Zhu, H., Zhang, M., Zhang, X., and Zhang, L. (2021). Two-branch encoding and iterative attention decoding network for semantic segmentation. *Neural Comput. Appl.* 33, 5151–5166. doi:10.1007/s00521-020-05312-9