SANet: A Sea–Land Segmentation Network Via Adaptive Multiscale Feature Learning

Binge Cui, Wei Jing, Ling Huang, Zhongrui Li, and Yan Lu

***Abstract*—Sea–land segmentation of remote sensing images is of great significance to the dynamic monitoring of coastlines. However, the types of objects in the coastal zone are complex, and their spectra, textures, shapes, and distribution features are different. Therefore, sea–land segmentation for various types of coastlines is still a challenging task. In this article, a scale-adaptive semantic segmentation network, called SANet, is proposed for sea–land segmentation of remote sensing images. SANet has made two innovations on the basis of the classic encoder–decoder structure. First, to integrate the spectral, textural, and semantic features of ground objects at different scales, we designed an adaptive multiscale feature learning module (AML) to replace the conventional serial convolution operation. The AML module mainly contains a multiscale feature extraction unit and an adaptive feature fusion unit. The former can capture the multiscale detailed information and contextual semantic information of objects from an early stage, while the latter can adaptively fuse feature maps of different scales. Second, we adopted the squeeze-and-excitation module to bridge the corresponding layers of the codec so that SANet can selectively emphasize the features of the weak sea–land boundaries. Experiments on a set of Gaofen-1 remote sensing images demonstrated that SANet achieved more accurate segmentation results and obtained sharper boundaries than other methods for various natural and artificial coastlines.**

***Index Terms*—Adaptive learning, atrous convolution, remote sensing image, residual block, sea–land segmentation, squeeze-excitation module.**

I. Introduction

SEA–LAND segmentation aims to separate coastal remote sensing images into ocean and land regions, which is a key step for many coastal applications, such as coastline change analysis [1], ship detection [2], and maritime safety [3]. A large number of automatic sea–land segmentation methods have been put forward, which can be mainly divided into the following categories: Thresholding segmentation methods, object-oriented segmentation methods, and methods based on machine learning or deep learning [4]–[13].

The thresholding segmentation methods are simple and easy to implement. This class includes the most commonly used methods in the sea–land segmentation of remote sensing images [4]–[6]. In a normalized difference image, pixels with intensities below the thresholding are classified as negative (land), and those with intensities above the thresholding are classified as positive (water). The thresholding segmentation methods obtain satisfactory results for coastal areas (such as artificial coasts), with large spectral differences between sea and land. However, the conventional thresholding segmentation methods only rely on the spectral information, and it is difficult to correctly distinguish objects with a similar spectrum, such as aquaculture ponds and the sea. Moreover, the selected optimal thresholding is easily affected by factors such as coast type, sensor, weather, and season, which limits the application of the thresholding method in complex sea–land segmentation scenarios. The object-oriented segmentation methods divide the image into objects of different sizes composed by image segmentation [7]–[9]. This method takes the object as the basic unit, ignores the textural features of the object, and processes the image according to the spectral and spatial features [10]. The object-oriented segmentation methods can reduce the interference of the internal information of the pixel, but their steps are complex, and they cannot make full use of the hidden information of the image. Machine learning can extract useful information and knowledge from a large amount of incomplete random data [10]. Recently, machine learning-based methods have also been applied to sea–land segmentation tasks [11]–[13]. The sea–land segmentation methods based on machine learning can achieve high automation, but it requires a combination of multiple machine learning methods to obtain better extraction results. Deep learning based on fully convolutional neural network has achieved satisfactory performance in the field of semantic segmentation [14]–[19]. Fully convolutional network can automatically extract the features from input images and reconstruct the image resolution through the decoder [14]. In recent years, work on sea–land segmentation of remote sensing images based on deep learning has also made great progress [20]–[22]. For example, DeepUNet introduced the residual block on the basis of U-Net to extend the depth of the network, thereby extracting deeper features for sea–land segmentation [20]. RDU-Net used dense connection blocks to enhance feature reuse [21]. SeNet added edge supervision to the structure of the deep semantic segmentation network, thereby obtaining more accurate sea–land boundaries [22].

The above methods exhibit high recognition and extraction accuracy for sea–land segmentation within a certain area. They demonstrate excellent performance in applications such as obtaining the location and length of the coastline in a small area. However, in actual applications, it is often necessary to investigate and analyze all coastlines of the entire region. With

Manuscript received August 28, 2020; revised October 30, 2020; accepted November 18, 2020. Date of publication November 24, 2020; date of current version January 6, 2021. This work was supported in part by the National Key R&D Program of China under Grant 2017YFC1405600 and in part by the National Natural Science Foundation of China (NSFC) under Grant 41 406 200. *(Corresponding author: Yan Lu.)*

The authors are with the College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266 590, China (e-mail: cuibinge@qq.com; wei_adam@126.com; sdythl@126.com; 1529494116@qq.com; luyan@sdust.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3040176

the expansion of the application area, some new problems also arise. In general, the following two problems exist in large-scale sea–land segmentation: 1) The scene is more complex, and various types of coastlines coexist. For example, the boundary of a port changes drastically, the breakwater extends deep into the sea, and the shape is long and narrow. These issues often coexist. The silty coast is open and flat, with a wide range of distribution, and changes slowly. It often accumulates with ground objects such as aquaculture ponds and estuaries. The spectral features of reclamation are close to those of silty or sandy coasts, and the latter will be classified as sea in the task of sea–land segmentation. (2) Weak boundaries (such as silt coastlines) and strong boundaries (such as artificial coastlines) are alternately distributed. It is difficult for the model to choose an optimal threshold or hyperparameter to determine the location of the sea–land boundary. Due to the existence of the above problems, most existing methods cannot obtain good segmentation results for large-scale sea–land segmentation tasks. In response to the above problems, a novel deep convolutional neural network (DCNN) model, called the scale-adaptive semantic segmentation network (SANet), is proposed in this article for sea–land segmentation. The SANet uses the designed adaptive multiscale learning (AML) module for multiscale feature extraction and fusion, which can obtain multiscale features of ground objects at the same resolution and adaptively learn the weight of each feature map based on the fused information. Moreover, SANet adopts the squeeze-and-excitation (SE) module to enhance the feature maps related to the weak sea–land boundaries then transfers them to the decoder to obtain a sharper segmentation boundary. To evaluate the performance of SANet for sea–land segmentation, experiments on a set of remote sensing image data from Gaofen-1 were carried out. Compared with other sea–land segmentation methods, SANet achieves higher accuracy, precision, recall, and F1-score values for both sea and land regions. The main contributions of this work are summarized as follows.

1) We designed a novel feature extraction and fusion module to replace the conventional serial convolution operation that can construct adaptive and multiscale contextual representations for sea–land segmentation tasks. The experimental results show that the AML module can be ported to other semantic segmentation models and significantly improve performance.

2) A novel DCNN model, called SANet, is proposed for sea–land segmentation of remote sensing images with complex coastline types. The SANet combines AML and SE modules to learn multiscale features and strengthen weak sea–land boundary information. Compared with other sea–land segmentation or semantic segmentation methods, SANet produces less misclassification, especially near the coastlines.

3) We provide the research community with a new high-quality dataset to advance sea–land segmentation with high-resolution remote sensing images. The dataset contains 1726 hand-labeled and cropped Gaofen-1 images with an 8-m spatial resolution and 4 bands, covering the various types of coastlines in Lianyungang, China. It can be found.[1]

[1][Online]. Available:https://www.kaggle.com/cuilab224/sea-land-segmentation-with-gaofen-1

II. Related Work

A. Sea–Land Segmentation

Most of the existing sea–land segmentation works are based on thresholding methods, e.g., the normalized difference water index (NDWI) [4], the modified normalized difference water index (MNDWI) [5], etc. Li proposed the second modified normalized difference water index (SMNDWI) to extract the waterline [23]. Liu proposed an automatic sea–land segmentation algorithm based on the locally adaptive thresholding technique [24]. You proposed a segmentation scheme, which can determine the threshold according to the adaptively established statistical model of the sea [25]. Chen proposed a threshold segmentation algorithm combining a rough threshold with an accurate threshold and provided a complete sea–land segmentation scheme [26]. The object-oriented methods are also applied to sea–land segmentation. Zhao *et al.* used an object-oriented segmentation method to automatically extract a wide range of water lines and classified the extracted coastlines based on the remote sensing interpretation symbols of different coastal types [8]. Lei *et al.* interpreted the sea–land segmentation task in view of superpixels, where similar pixels are clustered and the local similarity is explored [13].

Currently, the deep semantic segmentation network has been improved to make it suitable for sea–land segmentation tasks [20]–[22]. Li *et al.* [20] proposed a DeepUNet network based on U-Net structure. The network has two kinds of short connections, namely, the U connection and Plus connection. DeepUNet concatenates the layers in the encoder into the layers in the decoder and deepens U-Net to a deeper depth [20]. The network has achieved excellent segmentation results on natural color images from Google Earth. Shamsolmoali *et al.* [21] proposed a residual dense U-Net (RDU-Net) for pixelwise sea–land segmentation in complex and high-density remote sensing images. RDU-Net is a combination of both downsampling and upsampling paths and achieves satisfactory results [21]. Cheng *et al.* [22] proposed a local smooth regularization method for sea–land segmentation tasks to achieve better spatially consistent results, and used a multitask loss to simultaneously obtain the segmentation and edge detection results. The attached structured edge detection branch can further refine the segmentation results and dramatically improve the edge detection accuracy [22]. The abovementioned methods yield excellent performance in small-scale land and sea segmentation. However, when these methods are used for large-scale segmentation tasks, they show certain limitations in generalization ability.

B. Semantic Segmentation

In recent years, deep learning has achieved advanced performance in image segmentation, classification, and object detection [14]–[17]. Semantic segmentation models based on deep learning, such as the fully convolutional network (FCN) [14], U-Net [15], SegNet [16], and DeepLabv3+[17], have been proposed. Among them, FCN [14] replaces the fully connected layer in conventional convolutional neural networks (CNNs) with a convolution layer to classify images at the pixel level, performing end-to-end image segmentation tasks. U-Net [15] uses skip connections to concatenate the feature maps

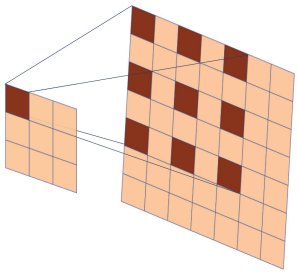


Fig. 1. Atrous Convolution with 3 × 3 kernel size and the dilation rate is 2.

of different scales generated in the encoder and corresponding feature maps in the decoder, which effectively preserves the boundary information. For CNNs, the more network layers there are, the richer the extracted features at different levels will be [27]–[30]. However, simply increasing the number of network layers will lead to gradient vanishing or gradient explosion and degradation [31], [32]. He K. *et al.* proposed a residual block to alleviate this problem [33]. By adding shortcut connections to the residual block, the network becomes easier to optimize. Based on the idea of residuals, the network can reach deeper depths and extract deeper levels of information.

For the semantic segmentation model, the larger the receptive field of the model is, the more contextual semantic information it can extract [30]. However, when increasing the receptive field by increasing the size of the kernel, the parameters of the model will increase significantly, and the model will be difficult to train. The atrous convolution was proposed to alleviate this problem [34]–[36]. This operation injects holes into the standard convolution map to increase the reception field. Compared with the general convolution operation, the atrous convolution has an additional hyperparameter called the dilation rate, which refers to the interval size of the kernel. For example, in Fig. 1, the receptive field of the standard 3 × 3 convolution kernel is expanded to 5 × 5 by atrous convolution with a dilation rate of 2, but the number of parameters does not change. The atrous convolution operation is as follows:

$$\mathrm{O}_{i,j}^{(d)} = \sigma\left(\sum_{l=0}^{k-1}\sum_{m=0}^{k-1} \boldsymbol{W}_{(l,m)} \cdot \mathrm{I}_{(d*l+i,d*m+j)} + \boldsymbol{b}\right) \quad (1)$$

where O refers to the output feature maps, d is the dilation rate, i, j refer to the indexes of the pixel, σ refers to the activation function, $\boldsymbol{W}$ refers to the weight, l, m refer to the parameter indexes of the convolution kernel, and k is the kernel size, I refers to the input feature maps, and $\boldsymbol{b}$ is the bias.

III. Proposed Method

The proposed sea–land segmentation model, which is called SANet, is introduced in this section. The structure of SANet is first demonstrated in Section III-A, and then the innovative AML module is presented in detail in Section III-B. Finally, we introduce the enhancement mechanism of the weak sea–land boundaries in SANet in Section III-C.

A. Overall Structure of SANet

We developed an end-to-end sea–land segmentation model. As shown in Fig. 2, the input is a four-band remote sensing image, and the output is a binary segmentation map in which 1 (blue pixels) represents sea and 0 (brown pixels) represents land. SANet retains U-Net's U-codec structure and skips connections. However, SANet uses the proposed AML module to perform feature extraction and fusion instead of the ordinary serial convolution operations. The AML module can provide multiscale receptive fields and adaptively adjust the weight of the feature maps. Moreover, SANet adopts the SE modules to bridge the corresponding layers of the codec. The SE module can enhance useful features to obtain a sharper sea–land boundary.

B. Adaptive Multiscale Feature Learning Module

Inspired by the atrous convolution operation [35] and attention mechanism [37], we propose an AML module to replace the conventional serial convolution operations. The AML module contains a multiscale feature extraction unit and an adaptive feature fusion unit. In the feature extraction unit, a residual branch and three atrous convolution branches with different dilation rates are designed, and these four branches can work in parallel to simultaneously extract detailed and multiscale contextual information. The four branches can provide 3 × 3, 5 × 5, 7 × 7, and 11 × 11 receptive field sizes. The feature fusion unit uses two branches to generate a learnable weight vector and joint feature maps and then fuses them by channelwise multiplication and the convolution operation with a 1 × 1 kernel. Next, we will specifically introduce these two units.

As shown in Fig. 3, a convolution operation with a 1 × 1 kernel is first performed on the input feature map I to generate the feature map F with the number of channels p as follows:

$$\mathrm{F} = \delta\left(\boldsymbol{W}_{1\times1} * \mathrm{I}\right) \quad (2)$$

where δ refers to the ReLU activation function, $*$ refers to the convolution operation, and W is the weight of the current convolution kernel. Through the above convolution operation, the number of feature-map channels is adjusted to p. The feature map F is then processed in parallel by a residual block and three atrous convolutions to generate feature maps at four scales, i.e., $\mathrm{R}, \mathrm{A}^{(2)}, \mathrm{A}^{(3)}$, and $\mathrm{A}^{(5)}$. The residual block consists of a convolution operation with a 3 × 3 kernel size and a convolution operation with a 1 × 1 kernel, which can be expressed as

$$\mathrm{R} = \mathrm{F} + \delta\left(\boldsymbol{W}_{1\times1} * \delta\left(\boldsymbol{W}_{3\times3} * \delta\left(\boldsymbol{W}_{3\times3} * \mathrm{F}\right)\right)\right) \quad (3)$$

where δ refers to the ReLU activation function, $*$ refers to the convolution operation, and W is the weight of the current convolution kernel. To avoid the grid effect [38] and maintain the continuity of information, the dilation rates of the three atrous convolutions are set to 2, 3, and 5. This can be expressed as

$$\mathrm{A}_{x,y}^{(d)} = \delta\left(\sum_{i=0}^{2}\sum_{j=0}^{2} \boldsymbol{W}_{(i,j)} \cdot \mathrm{F}_{(x+i*d,y+j*d)} + b\right) \quad (4)$$

where x, y refer to the pixel indexes of output feature map, δ refers to the ReLU activation function, i, j refer to the parameter indexes of the convolution kernel, d refers to the dilation rate, and b refers to the bias. After the multiscale feature extraction,

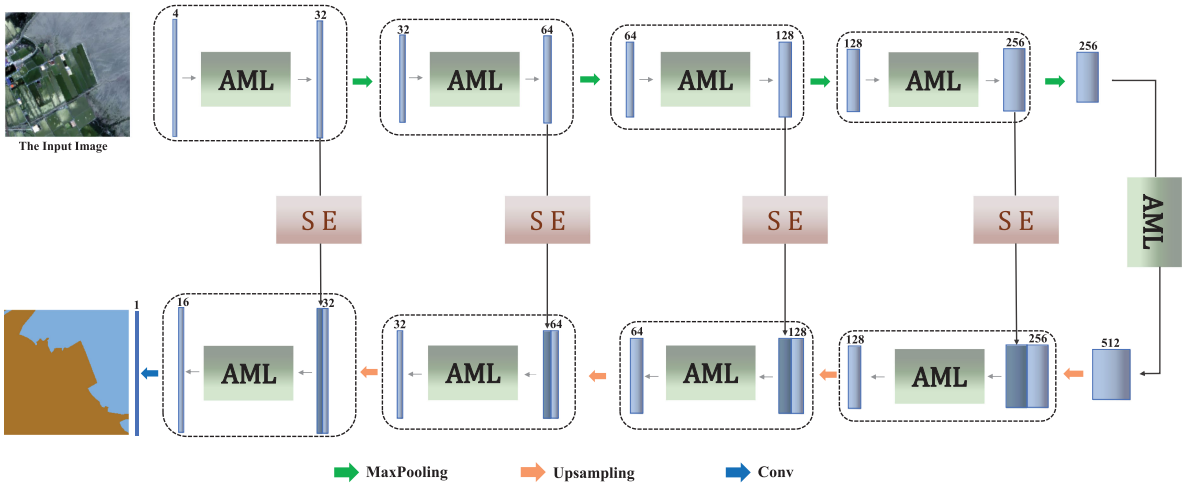


Fig. 2. Proposed SANet for sea–land segmentation.

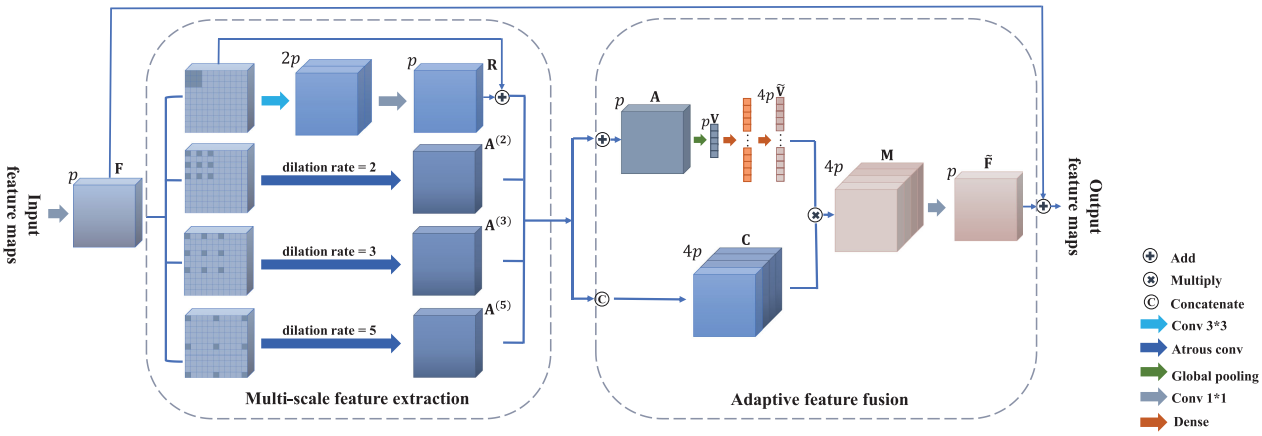


Fig. 3. Structure of AML module.

the feature maps generated by (5) and (6), respectively, added and concatenated, which can be expressed as

$$\mathrm{A} = \mathrm{R} + \mathrm{A}^{(2)} + \mathrm{A}^{(3)} + \mathrm{A}^{(5)} \tag{5}$$

$$\mathrm{C} = \delta\left(\boldsymbol{W}_{1\times1} * \left[\mathrm{R}\left|\left|\mathrm{A}^{(2)}\right|\right|\mathrm{A}^{(3)}||\mathrm{A}^{(5)}\right]\right) \tag{6}$$

where δ refers to the ReLU activation function, W is the weight of the current convolution kernel, and $||$ refers to concatenation operation. After that, the feature map A is squeezed into a p-dimensional vector V by the global average pooling operation to generate a channelwise global feature, and then V is trained by two dense layers to generate $4p$-dimensional vector $\tilde{\mathrm{V}}$, which can be expressed as

$$\tilde{\mathrm{V}} = \sigma\left(\delta\left(\boldsymbol{W}_1, \mathrm{V}\right), \boldsymbol{W}_2\right) \tag{7}$$

where σ refers to sigmoid activation function, δ refers to the ReLU activation function, and W is the weight of the current convolution kernel. Through channelwise multiplication, vector $\tilde{\mathrm{V}}$ is assigned to feature map C, and feature maps of different scales are fused by a convolution operation with 1×1 kernel size, which can be expressed as

$$\tilde{\mathrm{F}} = \delta\left(\boldsymbol{W}_{1\times1} * \left(\mathrm{C} \circ \tilde{\mathrm{V}}\right)\right) \tag{8}$$

where δ refers to the ReLU activation function, W is the weight of the current convolution kernel, and $\circ$ refers to the channelwise multiplication. Finally, F and $\tilde{\mathrm{F}}$ are added to generate output feature map O, thereby accelerating model learning and alleviating gradient disappearance.

C. Sea–Land Boundary Feature Enhancement

In large-scale remote sensing images, there are usually some weak sea–land boundaries, such as estuaries and silty coasts. To address the problem of weak boundaries, through experiments, we repeatedly observed the feature maps of various weak boundaries output by the AML module. We found that some feature maps have high response near the weak boundaries, while others have low response near the weak boundaries. It is obvious that the former contributes more to the recognition of sea–land boundaries. Enhancing the feature maps with high response to weak boundaries is very important for accurately

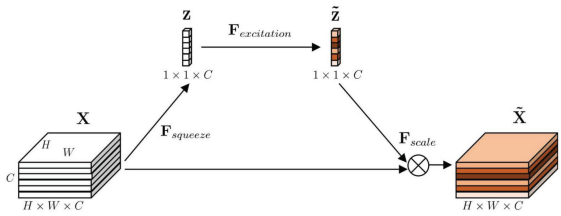


Fig. 4. Squeeze-and-excitation module.

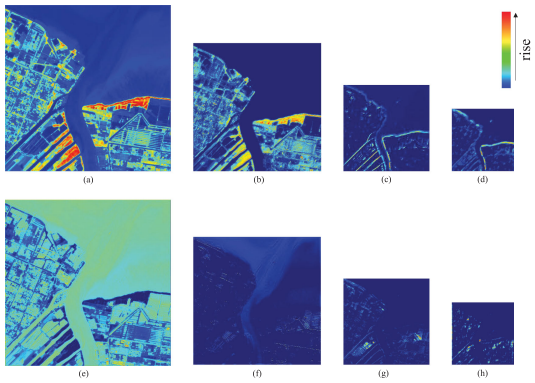


Fig. 5. Feature maps output by different layers with different weights. l refers to the output layer index and w refers to the weight assigned to the feature map by the SE module.

identifying the sea–land boundaries. In this article, we added the SE module between the corresponding layers of the codec to enhance the representation of weak boundaries.The structure of the SE module is shown in Fig. 4. Given the input feature map $X \in R_{H\times W\times C}$, where H, W, and C refer to the height, width, and the number of channels of feature maps, respectively. The SE module operates as follows: First, global average pooling [39], [40] is performed on each feature map to obtain the output $C-$ dimensional column vector z

$$z_k = F_{\text{squeeze}}\left(X_k\right) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_k(i, j) \tag{9}$$

where X_k refers to the kth feature map of X. Second, the output z undergoes excitation operation containing two fully connected layer and two activation functions to generate $\widetilde{z}$

$$\tilde{z} = F_{\text{excitation}}\left(z, \boldsymbol{W}\right) = \sigma\left(\delta\left(\boldsymbol{W}_1, z\right), \boldsymbol{W}_2\right) \tag{10}$$

where σ refers to sigmoid activation function, and δ refers to ReLU activation function. At last the F_{scale} operation rescales the given feature maps X with $\widetilde{z}$

$$\tilde{X}_k = F_{\text{scale}}\left(X_k, \tilde{z}_k\right) = X_k \circ \tilde{z}_k \tag{11}$$

where $X_k \circ \tilde{z}_k$, $\{k|k = 1, 2, 3 \cdots C\}$ refers to channelwise multiplication. Through the above operations, the SE module selectively enhances the feature maps with high response near weak boundaries and suppresses those that have low response near weak boundaries. We visualized some feature maps at each layer. As shown in Fig. 5, the SE module assigns high weights to feature maps with a higher response near the weak boundaries in the first row, and assigns low weights to feature maps with a lower response near the weak boundaries in the second row.

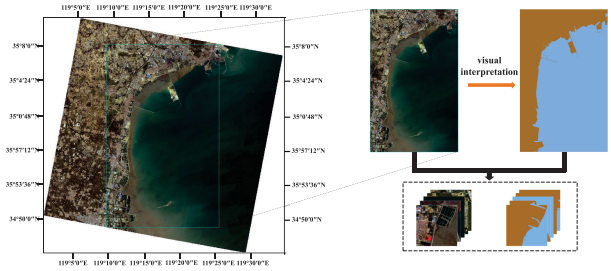


Fig. 6. Illustration of data preparation process.

IV. Experiments and Evaluation

In this section, we construct a dataset to test the performance and robustness of SANet. The dataset is introduced in Section IV-A. Section IV-B introduces details of the experiment. Section IV-C presents the experimental results.

A. Experimental Data Preparation

The experimental dataset was acquired from nine multispectral remote sensing images shot by Gaofen-1 in the Lianyungang coastal zone, Jiangsu Province, China. The Gaofen-1 images contain four bands, namely, red, green, blue and near-infrared bands, and the spatial resolution is 8 m. Each selected image contains a different type of coastline. As shown in Fig. 6, the remote sensing image around the coast is cropped and labeled by experts through visual interpretation. The ground truth map is a binary image, with 0 representing sea and 255 representing land. The labeled image is divided into 256 $\times$ 256 samples by checkerboard segmentation. The training set contains 1544 samples, the validation set contains 178 samples, and the test set contains 192 samples.

B. Implementation Details

Setup: The experiment was conducted on a server equipped with NVIDIA Tesla P100 GPU with 16 GB of graphics memory, and CentOS Linux 7.5 operating system. All models were trained and tested with Keras framework [41], using TensorFlow as the backend engine [42]. During training process, the Adam algorithm was used to minimize the loss, and we set the initial learning rate to 0.0001 and the number of epochs to 100.

Metrics: To evaluate our models, four metrics (accuracy rate, precision rate, recall rate, and F1-score) were used to assess the experimental results. The metrics are defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{12}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{13}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{14}$$

$$\text{F}_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{15}$$

where TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively.

TABLE I
EVALUATION RESULTS OF DIFFERENT METHODS ON THE TEST SET

Methods	Accuracy(%)	Precision(%)	Recall(%)	F1-score
NDWI	80.13	79.79	77.31	0.7853
Multiresolution	93.03	98.84	91.26	0.9249
SVM	89.74	87.03	91.86	0.8938
U-Net	94.69	91.96	97.19	0.9450
SegNet	95.21	93.95	95.98	0.9496
DeepLabv3+	95.02	94.04	95.46	0.9475
DeepUNet	95.89	95.61	95.66	0.9563
SANet	**98.63**	**98.44**	**98.65**	**0.9855**

In each row, the number in bold is the largest value.

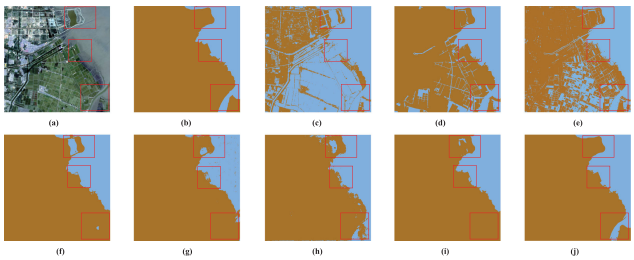


Fig. 7. Sea–land segmentation on test image 1. (a) Test image. (b) Ground truth. (c) NDWI. (d) Multiresolution-segmentation. (e) SVM. (f) U-Net. (g) SegNet. (h) DeepLabv3+. (i) DeepUNet. (j) SANet. The number 0.14 is the best thresholding obtained in the experiment.

C. Experimental Results

We compared the thresholding segmentation method (NDWI [4]), object-oriented segmentation method (Multiresolution-segmentation [8], [9]), support vector machine (SVM [11]), U-Net [15], SegNet [16], DeepLabv3+[17], DeepUNet [20], and the proposed SANet in the same experimental environment. Table I shows the quantitative results of the above methods on the test set. The NDWI thresholding segmentation method has the lowest accuracy. SANet's accuracy is 5.60% higher than that of multiresolution-segmentation, 8.89% higher than that of the SVM, 3.94% higher than that of U-Net, 3.42% higher than that of SegNet, 3.61% higher than that of DeepLabv3+, and 2.74% higher than that of DeepUNet. SANet's F1-score is 0.0606, 0.0917, 0.0405, 0.0359, 0.0380, and 0.0292 higher than those of multiresolution-segmentation, SVM, U-Net, SegNet, DeepLabv3+, and DeepUNet, respectively. Four representative remote sensing images containing different types of coastlines were shown to evaluate the performance of each method. The sea–land segmentation results are shown in Figs. 7–14.

Test image 1 is a coastal zone containing a large number of aquaculture ponds and silt. In Fig. 7(c) and (e), because aquaculture ponds and seawater have similar spectral features, and shoals composed of silt are similar to land in terms of the spectral features, the NDWI and SVM methods classify aquaculture ponds as sea and some of the shoals as land. In Fig. 7(d), the multiresolution-segmentation method identifies slender-scale canals and aquaculture fences as sea. In Fig. 7(g) and (h), the extracted land boundary is not well aligned with the boundary of the aquaculture pond, which may be because the detailed information is not directly propagated to the decoders in SegNet and DeepLabv3+. In the bottom rectangle of Fig. 7, the results of various models for identifying the estuary with a large amount of suspended sediment are poor, and SANet performs better than the other models. Fig. 8 shows the details of the segmentation near the ring levee. In Fig. 8(c) and (f), NDWI threshold segmentation and U-Net completely identify the water area inside the ring levee as the sea. SegNet, DeepLabv3+, and DeepUNet yield partial misclassification in this area. The evaluation results on test image 1 are listed in Table II. The indicators show that SANet's accuracy is 5.72% higher than that of multiresolution-segmentation, 7.47% higher than that of the SVM, 7.08% higher than that of U-Net, 6.29% higher than that of SegNet, 7.73% higher than that of DeepLabv3+, and 6.35% higher than that of DeepUNet. SANet's F1-score is 0.0581 higher than that of multiresolution-segmentation, 0.0772 higher than that of the SVM, 0.0649 higher than that of U-Net, 0.0559 higher than that of SegNet, 0.0731 higher than that of DeepLabv3+, and 0.0560 higher than that of DeepUNet.

Test image 2 is an artificial coast with a port and a large number of breakwaters. In Fig. 9(c) and (e), the NDWI and SVM methods based on single-pixel spectral information achieve good results for identifying breakwaters due to large differences in spectral information between sea and land. For slender breakwaters, due

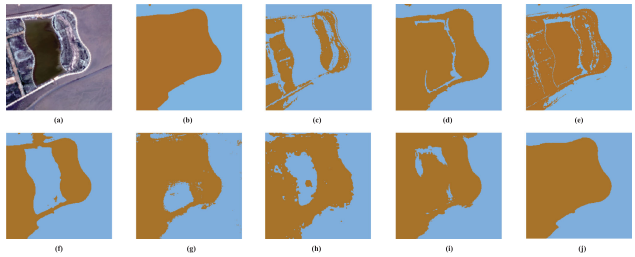


Fig. 8. Segmentation details on area 1. (a) Test image. (b) Ground truth. (c) NDWI. (d) Multiresolution-segmentation. (e) SVM. (f) U-Net. (g) SegNet. (h) DeepLabv3+. (i) DeepUNet. (j) SANet.

TABLE II
EVALUATION RESULTS ON TEST IMAGE 1

Methods	Accuracy(%)	Precision(%)	Recall(%)	F1-score
NDWI	72.59	**99.75**	48.54	0.6530
Multiresolution	93.03	99.55	87.27	0.9301
SVM	91.28	99.52	83.99	0.9110
U-Net	91.67	90.39	94.36	0.9233
SegNet	92.46	89.18	97.66	0.9323
Deeplabv3+	91.02	91.89	91.13	0.9151
DeepUNet	92.40	88.65	98.27	0.9322
SANet	**98.75**	98.71	**98.93**	**0.9882**

In each row, the number in bold is the largest value

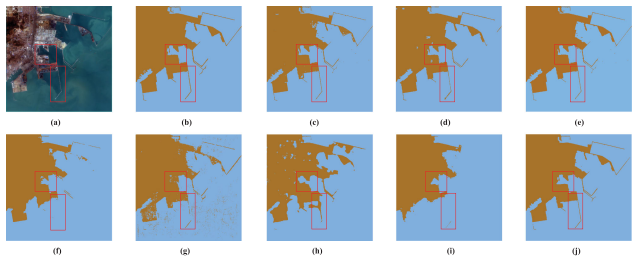


Fig. 9. Sea–land segmentation on test image 2. (a) Test image. (b) Ground truth. (c) NDWI. (d) Multiresolution-segmentation. (e) SVM. (f) U-Net. (g) SegNet. (h) DeepLabv3+. (i) DeepUNet. (j) SANet. The number 0.45 is the best thresholding obtained in the experiment.

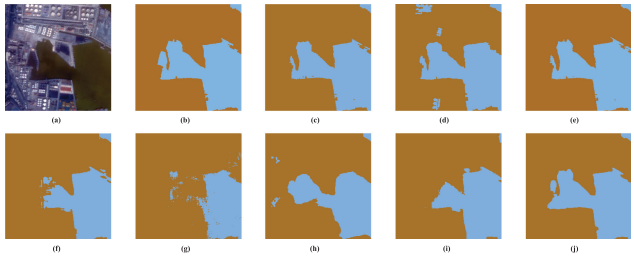


Fig. 10. Segmentation details on area 2. (a) Test image. (b) Ground truth. (c) NDWI. (d) Multiresolution-segmentation. (e) SVM. (f) U-Net. (g) SegNet. (h) DeepLabv3+. (i) DeepUNet. (j) SANet.

TABLE III
EVALUATION RESULTS ON TEST IMAGE 2

Methods	Accuracy(%)	Precision(%)	Recall(%)	F1-score
NDWI	98.79	98.52	**99.56**	**0.9904**
Multiresolution	98.49	97.94	96.51	0.9722
SVM	**98.93**	98.38	97.69	0.9804
U-Net	95.03	94.33	87.09	0.9056
SegNet	94.68	88.43	92.68	0.9051
DeepLabv3+	93.94	85.25	94.15	0.8948
DeepUNet	91.67	91.96	76.24	0.8336
SANet	98.86	**98.92**	96.91	0.9790

In each row, the number in bold is the largest value.

to the loss of detailed information caused by multiple downsampling, most deep learning-based methods yield poor recognition results. In Fig. 9(d), the multiresolution-segmentation method misclassifies small objects on land. In Fig. 9(a), the area marked by the middle rectangle is a port, and its enlarged snapshot is shown in Fig. 10. NDWI, multiresolution-segmentation, and SVM methods exhibit better results on the sea–land boundaries. SegNet misidentifies the shadow in the sea as land. U-Net, DeepLabv3+, and DeepUNet segment the port incompletely. SANet completely segments the entire port, but the edges are relatively smoother than the ground truth. The spectral features of the breakwater in the bottom rectangle are greatly affected by seawater due to its slender structure. U-Net and DeepUNet cannot recognize the slender structure of the breakwater due to the small receptive field during the early low-level information extraction process. SegNet and DeepLabv3+ lose the boundary information of the breakwater, so cracked or fuzzy breakwaters appear in the recognition results. SANet has dense-scale receptive fields and retains boundary information well, so it can identify most breakwaters. For test image 2, the evaluation results are listed in Table III. The SVM method has the highest accuracy. Compared with that of the SVM method, the accuracy of SANet is only 0.07% lower. Compared with other deep learning methods, SANet's accuracy is 0.37, 3.83, 4.18, 4.92, and 7.19 percentage points higher than those of multiresolution-segmentation, U-Net, SegNet, DeepLabv3+, and DeepUNet, respectively. SANet's F1-score is 0.0114 lower than that of NDWI and 0.0068, 0.0734, 0.0739, 0.0842, and 0.1454 higher than those of multiresolution-segmentation, U-Net, SegNet, DeepLabv3+, and DeepUNet, respectively.

Test image 3 is an artificial coast that contains some reclamations and breakwaters. In Fig. 11(a), the rectangle is the reclamation area, the spectral information of the reclamation area is similar to that of silty tidal flats, and the tidal flats were labeled

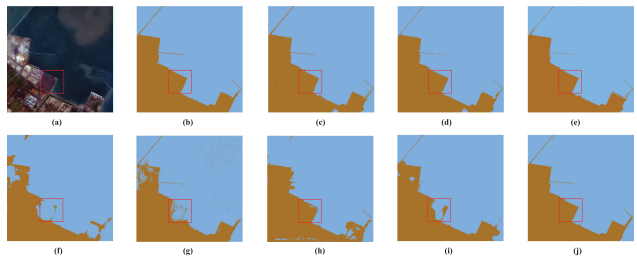


Fig. 11. Sea–land segmentation on test image 3. (a) Test image. (b) Ground truth. (c) NDWI. (d) Multiresolution-segmentation. (e) SVM. (f) U-Net. (g) SegNet. (h) DeepLabv3+. (i) DeepUNet. (j) SANet. The number 0.50 is the best thresholding obtained in the experiment.

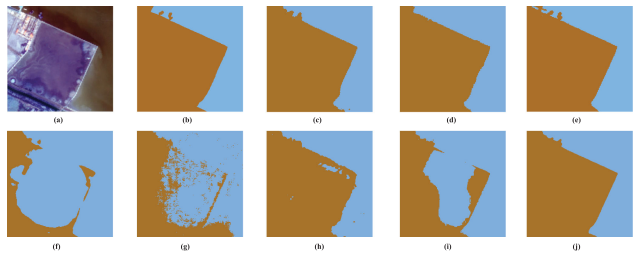


Fig. 12. Segmentation details on area 3. (a) Test image. (b) Ground truth. (c) NDWI. (d) Multiresolution-segmentation. (e) SVM. (f) U-Net. (g) SegNet. (h) DeepLabv3+. (i) DeepUNet. (j) SANet.

TABLE IV
EVALUATION RESULTS ON TEST IMAGE 3

Methods	Accuracy(%)	Precision(%)	Recall(%)	F1-score
NDWI	98.27	**99.13**	97.82	0.9847
Multiresolution	98.59	98.84	97.89	0.9836
SVM	98.49	98.61	98.76	0.9861
U-Net	88.47	96.26	76.24	0.8509
SegNet	93.39	95.96	88.40	0.9202
DeepLabv3+	95.54	94.65	95.05	0.9485
DeepUNet	95.10	96.92	91.55	0.9416
SANet	**99.35**	98.99	**99.52**	**0.9925**

In each row, the number in bold is the largest value.

as sea in the training set. Fig. 12 shows segmentation details of the reclamation area. U-Net, SegNet, and DeepUNet lost the structural information of reclamation, which led to the identification of the reclamation area as sea. In Fig. 11(h), DeepLabv3+ can identify most reclamation areas, but the detailed information of the breakwater is lost. Compared with other methods, SANet can provide large receptive fields at an early stage to obtain the spatial structure information of large ground objects such as reclamation areas, which helps to avoid misclassification. For test image 3, the evaluation results are listed in Table IV. SANet's accuracy is 1.08, 0.76, 0.86, 10.88, 5.96, 5.81, and 4.25 percentage points higher than those of NDWI, multiresolution-segmentation, SVM, U-Net, SegNet, DeepLabv3+, and DeepUNet, respectively. SANet's F1-score is 0.0078, which is 0.0089, 0.0064, 0.1416, 0.0723, 0.0440, and 0.0509 higher than those of NDWI, multiresolution-segmentation, SVM, U-Net, SegNet, DeepLabv3+, and DeepUNet, respectively. The NDWI method has the highest precision, which is 0.14 percentage points higher than that of SANet, while the recall rate is 1.70 percentage points lower than that of SANet.

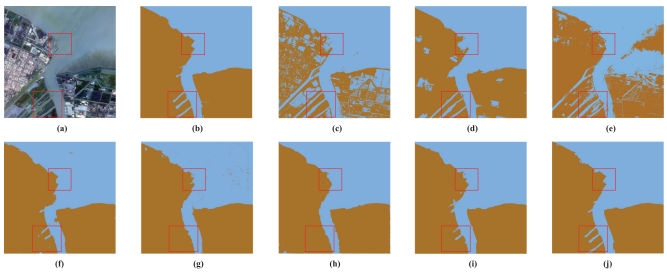


Fig. 13. Sea–land segmentation on test image 4. (a) Test image. (b) Ground truth. (c) NDWI. (d) Multiresolution-segmentation. (e) SVM. (f) U-Net. (g) SegNet. (h) DeepLabv3+. (i) DeepUNet. (j) SANet. The number 0.23 is the best thresholding obtained in the experiment.

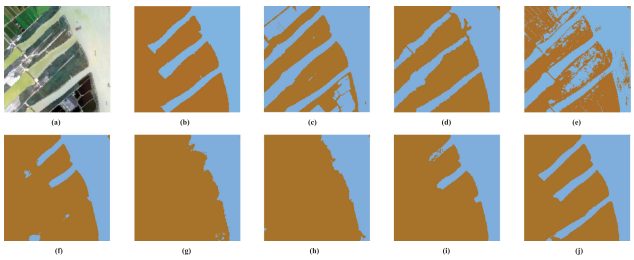


Fig. 14. Segmentation details on area 4. (a) Test image. (b) Ground truth. (c) NDWI. (d) Multiresolution-segmentation. (e) SVM. (f) U-Net. (g) SegNet; (h) DeepLabv3+. (i) DeepUNet. (j) SANet.

Test image 4 contains an estuary and biogenic coast. In Fig. 13(c), the NDWI method recognizes most aquaculture ponds as the sea. In Fig. 13, the multiresolution-segmentation method exhibits good classification for the sea part, but there are still some classification errors due to the complex scale of the inland objects. In Fig. 13(e), the SVM method misclassifies suspended sediment and aquaculture areas. In the upper rectangle of Fig. 13, DeepLabv3+ and SegNet cannot preserve the coastal boundary due to insufficient detailed information. In Fig. 13(a), the area marked by the bottom rectangle is an estuary that contains several rivers, and Fig. 14 shows its segmentation details. By convention, the part of the river between the estuary and the first bridge is defined as sea. Compared with other methods, SANet can better follow this convention to perform sea–land segmentation. NDWI, multiresolution-segmentation, and SVM methods identify the entire river as sea, and SegNet, DeepLabv3+, U-Net, and DeepUNet identify all or most of the rivers as land. At each layer of the model, there is a branch that obtains detailed information of small-scale ground objects and passes it to the decoder after being enhanced by the SE module. Therefore, it performs well on ground objects such as seaside estuaries. For test image 4, the evaluation results are listed in Table V. SANet's accuracy is 22.96, 6.24, 14.97, 2.50, 2.69, 3.20, and 2.12 percentage points higher than those of the NDWI method, multiresolution-segmentation, SVM, U-Net, SegNet, DeepLabv3+, and DeepUNet, respectively. SANet's F1-score is 0.2211 0.0512, 0.1294, 0.0164, 0.0193, 0.0247, and 0.0167 higher than those of the NDWI method, multiresolution-segmentation, SVM, U-Net, SegNet, DeepLabv3+, and DeepUNet, respectively.

TABLE V
EVALUATION RESULTS ON TEST IMAGE 4

Methods	Accuracy(%)	Precision(%)	Recall(%)	F1-score
NDWI	75.74	98.53	63.02	0.7687
Multiresolution	92.46	97.94	90.11	0.9386
SVM	83.73	95.40	78.53	0.8604
U-Net	96.54	95.84	98.89	0.9734
SegNet	96.20	96.51	97.59	0.9705
DeepLabv3+	95.50	95.81	97.21	0.9651
DeepUNet	96.58	97.86	96.78	0.9731
SANet	**98.70**	**99.08**	**98.88**	**0.9898**

In each row, the number in bold is the largest value

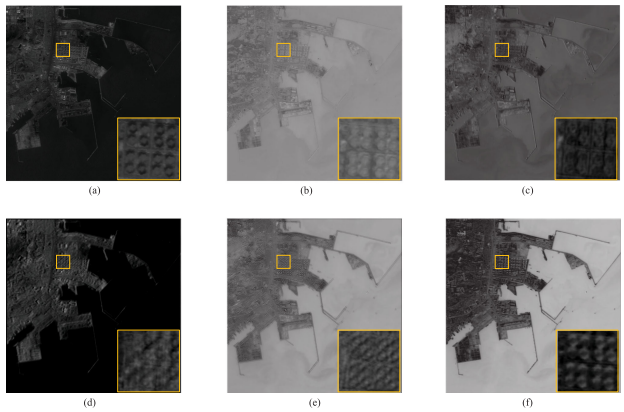


Fig. 15. Input feature map, feature maps of different receptive fields, and output feature map of the first AML module.

V. ANALYSIS AND DISCUSSION

A. Analysis

The Role of the AML Module: By providing multiscale receptive fields, the proposed AML module can extract and fuse multiscale features for sea–land segmentation. In Fig. 15, we visualized the input feature map, the feature map of each branch, and the output feature map of the first AML module. The locally enlarged area in Fig. 15 is an oil tank farm. In Fig. 15(b), the feature map output by the residual branch of the AML module contains rich detailed information. Due to the enlarged receptive field, some of the detailed information of the oil tank in Fig. 15(c) is lost. In Fig. 15(d), most of the detailed information of a single oil tank has been lost, and the textural features of the entire oil tank farm have become visible. In Fig. 15(e), the textural features of the entire tank farm are more obvious. In Fig. 15(f), the output feature map fuses feature maps of multiscale receptive fields, which contain both the detailed information of a single oil tank and the textural features of the entire oil tank farm. The AML module can extract and fuse multi-scale representations at each layer, so that SANet has more powerful feature extraction capabilities.

Ablation Study: To thoroughly investigate the effectiveness of the proposed method, we conduct ablation experiments by removing specific components for comparison. As shown in Table VI, the U-Net with AML module can achieve 3.35%, 2.07%, 9.75%, and 1.63% accuracy improvements on the four test images. The U-Net with the SE module can achieve 2.38%, 1.81%, 7.08%, and 0.78% accuracy improvements on the four

TABLE VI
ABLATION EVALUATION RESULTS

Testdata	Metrics	U-Net	U-Net+SE	U-Net+AML	SANet
Image 1	Accuracy(%)	91.67	94.05	95.02	**98.75**
	F1-score	0.9233	0.9435	0.9548	**0.9882**
Image 2	Accuracy(%)	95.03	96.84	97.10	**98.86**
	F1-score	0.9056	0.9401	0.9445	**0.9790**
Image 3	Accuracy(%)	88.47	95.55	98.22	**99.35**
	F1-score	0.8509	0.9478	0.9791	**0.9925**
Image 4	Accuracy(%)	96.54	97.32	98.17	**98.70**
	F1-score	0.9734	0.9791	0.9857	**0.9898**

In each row, the number in bold is the largest value.

TABLE VII
AML MODULE EVALUATION RESULTS

Testdata	Metrics	SegNet	SegNet +AML	DeepUNet	DeepUNet +AML
Image 1	Accuracy (%)	92.46	96.08	92.40	93.92
	F1-score	0.9323	0.9642	0.9322	0.9425
Image 2	Accuracy (%)	94.68	95.26	91.67	93.74
	F1-score	0.9056	0.9095	0.8336	0.8834
Image 3	Accuracy (%)	93.39	95.32	95.10	96.30
	F1-score	0.9202	0.945	0.9416	0.9560
Image 4	Accuracy (%)	96.20	97.66	96.58	96.70
	F1-score	0.9705	0.9819	0.9731	0.9751

test images. SANet combines the adaptive multiscale receptive fields of the AML module and the attention mechanism of the SE module, which can achieve 7.08%, 3.83%, 10.88%, and 2.16% accuracy improvements on the four test images. The above data fully illustrate the effectiveness and complementarity of the AML module and SE module in the sea–land segmentation task.

Portability of the AML Module: We use SegNet and DeepUNet as the baseline networks to evaluate the performance improvements contributed by the AML module. As shown in Table VII, the SegNet with AML module can achieve 3.62%, 0.58%, 1.93%, and 1.46% accuracy improvements on the four test images. The DeepUNet with AML module can achieve 1.52%, 2.07%, 1.20%, and 0.12% accuracy improvements on the four test images. The results prove that AML modules can be ported to other similar network architectures and contribute to significant performance improvements.

Hyperparameter Settings: Hyperparameter p is used to control the number of channels of the output feature map of each AML module in SANet. In SANet, the respective numbers of output channels of each AML module are p, $2p$, $4p$, $8p$, $16p$, $4p$, $2p$, p, and $0.5p$, respectively. We carried out experiments to reflect the influence of different p-values on the overall accuracy. As shown in Fig. 16, when p increases from 4 to 32, the overall accuracy is continuously improved. When $p = 32$, the overall accuracy reaches its peak. After that, as the value of p continues to increase, the overall accuracy exhibits a downward trend. Therefore, the optimal value of p is set to 32.

Using Skip Connections: The skip connections play an important role in SANet. Experiments were carried out to explore the influence of using skip connections on the overall accuracy. The experimental results in Fig. 17 show that the overall accuracy is significantly improved when skip connections are added to SANet. As the number of skip connections increases, the overall accuracy continues to improve, which shows the advantages of

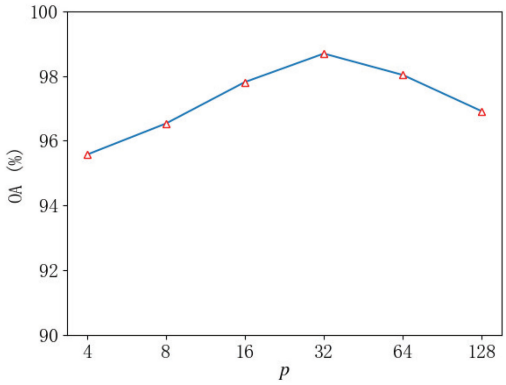


Fig. 16. Overall accuracy of SANet with different values of hyperparameter p.

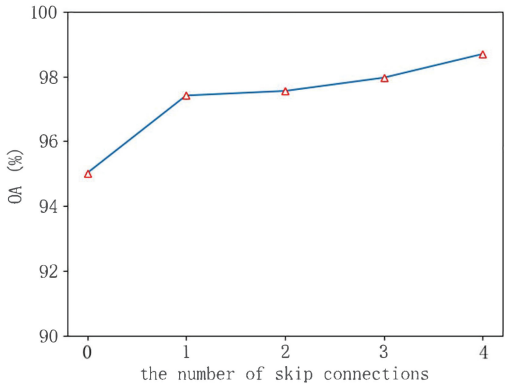


Fig. 17. Overall accuracy of SANet with different numbers of skip connections.

TABLE VIII
TIME CONSUMPTION AND PARAMETER OF DIFFERENT MODELS

Method	Train time(s/epoch)	Test time(s)	Parameters
U-Net	104	0.23	31.0M
SegNet	147	0.34	29.4M
DeepLabv3+	170	0.85	41.3M
DeepUNet	67	0.11	4.91M
SANet	149	0.55	27.3M

using multiple skip connections. Therefore, skip connections are added to each corresponding layer between the codecs of SANet.

Evaluation of Model Complexity: We compare the number of parameters and runtimes of the different models. As shown in Table VIII, SANet has more parameters than DeepUNet and fewer parameters than U-Net, SegNet and DeepLabv3+. Compared with those of other models, the training time of SANet for each epoch increases slightly. However, as shown in Fig. 18, SANet converges much faster than other models, which can compensate for the shortcoming that each epoch takes longer to run. In addition, in sea–land segmentation applications, we need to pay more attention to the runtime of the test process, and that of SANet is acceptable.

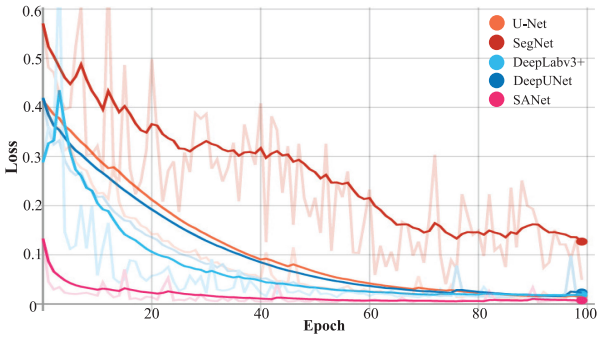


Fig. 18. Loss of different models on the verification set.

B. Discussion

Sea–land segmentation has significant implications for coastal zone management and coastal zone evolution research. Sea–land segmentation is more complicated than water–land segmentation. More semantic information must be combined to determine whether the surface corresponding to each pixel is land or sea. The thresholding segmentation methods only consider the spectral features of pixels, which take waterline extraction as the basis of sea–land segmentation [4]. However, the threshold of these kinds of methods is difficult to determine. The object-oriented segmentation methods can reduce the influence of the internal textural features of the ground objects on the segmentation results [8]. However, for a large-scale remote sensing image, the optimal segmentation scale is difficult to determine, and steps such as classification are required after object segmentation, making the process relatively complicated. The methods based on machine learning need to define spectral and spatial features in advance and to achieve good results, they often need to combine multiple machine learning strategies [10].

Deep learning provides new ideas for sea–land segmentation of large-scale remote sensing images. Researchers have improved the deep semantic segmentation architecture for natural images to make it suitable for sea–land segmentation tasks. DeepUNet extends the depth of U-Net to extract deeper context features and adopts the residual structure to avoid overfitting [20]. RDU-Net mainly considers feature reuse to make full use of hierarchical features in the original images [21]. However, these methods did not pay special attention to the coexistence of multiple types of coastlines in the same remote sensing image. When the shape, size, and distribution of various coastlines differ greatly, limitations are observed. In this article, considering the alternating distribution of various types of coastlines in large-scale practical applications, SANet implements adaptive learning of multiscale contextual semantic and detailed information. Moreover, SANet adopts the SE module to enhance weak sea–land boundary features. Therefore, SANet has more robust feature extraction capabilities and more powerful generalization capabilities, making it suitable for large-scale sea–land segmentation tasks. Considering that the spatial resolution of the images greatly influences the details of ground objects, in the future, we will try to use panchromatic and multispectral fusion images for sea–land segmentation. Moreover, based on the idea of integrated learning, we will design an edge optimization branch to guide the sea–land segmentation to further improve the accuracy of boundary extraction.

VI. Conclusion

In this article, we design a novel deep learning model, called SANet, for the sea–land segmentation of large-scale remote sensing images. SANet has the following attractive properties: 1) The proposed AML module can extract and adaptively fuse multiscale context representations, thereby improving the performance and adaptability of the model in large-scale practical applications in complex scenes. 2) By adopting the SE module, SANet can adaptively strengthen weak sea–land boundaries, so that the sea–land segmentation results have better spatial consistency. The above two advantages enable SANet to adapt to complex scenarios where various types of coastlines are alternately distributed. To verify the network architecture, we performed experiments on a set of Gaofen-1 remote sensing images containing different types of coastlines, and the experimental results show that the proposed SANet model significantly outperformed the other models.

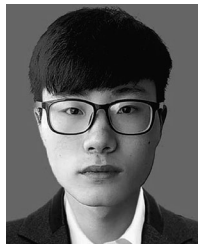
References

[1] W. Xiaojuan, X. Chenchao, C. Zhenying, and L. Xiaoji, "Coastline extraction based on object-oriented method using GF-2 satellite data," *Spacecraft Recovery Remote Sens.*, vol. 36, no. 4, pp. 84–92, 2015.
[2] X. Huihui, X. Qizhi, and H. Lei, "A sea-land segmentation algorithm based on gray smoothness ratio," in *Proc. 4th Int. Workshop Earth Observ. Remote Sens. Appl.*, 2016, pp. 117–121.
[3] X. Shi and L. Ge, "Method of water and land segmentation in optical remote sensing images," *Foreign Electron. Meas. Technol.*, 2014, pp. 29–32.
[4] S. K. McFeeters, "The use of the normalized difference water index (ndwi) in the delineation of open water features," *Int. J. Remote Sens.*, vol. 17, no. 7, pp. 1425–1432, 1996.
[5] M. H. A. Baig, L. Zhang, S. Wang, G. Jiang, S. Lu, and Q. Tong, "COmparison of MNDWI and DFI for water mapping in flooding season," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2013, pp. 2876–2879.
[6] L. ZHANG and Y. HU, "Improved roewa operator for sea-land segmentationin SAR image," *Comput. Eng. Appl.*, vol. 2017, no. 18, p. 26, 2017.
[7] Y. Chen, T. Feng, P. Shi, and J.-f. Wang, "Classification of remote sensing image based on object oriented and class rules," *Geomatics Inf. Sci. Wuhan Univ.*, vol. 31, no. 4, pp. 316–320, 2006.
[8] Z. Zhiling, L. I. Hui, D. Yuee, X. U. Wei, W. Ping, and J. Linhai, "Object-oriented waterline extraction based on gf-1 satellite remote sensing images," *Spacecraft Recovery Remote Sens.*, vol. 38, no. 4, 2017, pp. 106–116.
[9] M. Baatz and A. Schäpe, "Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation," in *Beitrge zum AGIT-Symp.*, 2000, pp. 12–23.
[10] T. Wei and L. Wensong, "Research on coastline automatic extraction methods based on remote sensing images," in *Proc. Int. Conf. Ind. Technol. Manage. Sci.*
[11] Y. Wang, Q. Yu, W. Lv, and W. Yu, "Coastline detection in SAR images using multi-feature and SVM," in *Proc. 4th Int. Congr. Image Signal Process.*, vol. 3, 2011, pp. 1227–1230.
[12] Z. Ming, Y. Bai-Long, H. E. Min, C. Zheng-Zheng, and Z. Xiong-Mei, "A sea-land segmentation of SAR image based on improved slic algorithm," *Electron. Opt. Control*, 2019, pp. 21–25.
[13] S. Lei, Z. Zou, D. Liu, Z. Xia, and Z. Shi, "Sea-land segmentation for infrared remote sensing images based on superpixels and multi-scale features," *Infrared Phys. Technol.*, vol. 91, pp. 12–17, 2018.
[14] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Nov. 2014.
[15] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," 2015.

[16] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[17] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," pp. 833–851, 2018.
[18] X. Xu, B. Pan, Z. Chen, Z. Shi, and T. Li, "Simultaneously multiobjective sparse unmixing and library pruning for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, to be published.
[19] J. Zhang, J. Liu, B. Pan, and Z. Shi, "Domain adaptation based on correlation subspace dynamic distribution alignment for remote sensing image scene classification," *IEEE Trans. Geoence Remote Sens.*, vol. PP, no. 99, pp. 1–11, Nov. 2020.
[20] R. Li, W. Liu, L. Yang, S. Sun, W. Hu, F. Zhang, and W. Li, "DeepuNet: A deep fully convolutional network for pixel-level sea-land segmentation," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 3954–3962, Nov. 2018.
[21] P. Shamsolmoali, M. Zareapoor, R. Wang, H. Zhou, and J. Yang, "A novel deep structure u-Net for sea-land segmentation in remote sensing images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. PP, no. 99, pp. 1–14, Sep. 2019.
[22] D. Cheng, G. Meng, G. Cheng, and C. Pan, "SeNet: Structured edge network for sea–land segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 2, pp. 247–251, Feb. 2017.
[23] M. Li and X. Zheng, "A second modified normalized difference water index (SMNDWI) in the case of extracting the shoreline," *Mar. Sci. Bull.*, vol. 18, no. 02, pp. 15–27, 2016.
[24] H. Liu and K. Jezek, "Automated extraction of coastline from satellite imagery by integrating canny edge detection and locally adaptive thresholding methods," *Int. J. Remote Sens.*, vol. 25, no. 5, pp. 937–958, 2004.
[25] X. You and W. Li, "A sea-land segmentation scheme based on statistical model of sea," in *Proc. 4th Int. Congr. Image Signal Process.*, vol. 3, 2011, pp. 1155–1159.
[26] X. Chen, J. Sun, K. Yin, and J. Yu, "Sea-land segmentation algorithm of SAR image based on Otsu method and statistical characteristic of sea area," *J. Date Acquis. Process*, vol. 29, pp. 603–608, 2014.
[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Neural Inf. Process. Syst.*, vol. 25, Jan. 2012, doi: 10.1145/3065386.
[28] L. Shi, Z. Wang, B. Pan, and Z. Shi, "An end-to-end network for remote sensing imagery semantic segmentation via joint pixel- and representation-level domain adaptation," *IEEE Geoence Remote Sens. Lett.*, vol. PP, no. 99, pp. 1–5, 2020.
[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014.
[30] C. Szegedy *et al.*, "Going deeper with convolutions," Sep. 2015, *arXiv:1409.4842*.
[31] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," 2015, *arXiv:1506.00019*.
[32] X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," pp. 802–810, 2015.
[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," pp. 770–778, 2016.
[34] L. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.
[35] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," May 2017, *arXiv:1705.09914*.
[36] B. Pan, X. Xu, Z. Shi, N. Zhang, and X. Lan, "DSSNET: A simple dilated semantic segmentation network for hyperspectral imagery classification," *IEEE Geosci. Remote Sens. Lett.*, vol. PP, no. 99, pp. 1–5, Nov. 2020.
[37] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 510–519.
[38] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," 2015, *arXiv:1506.00019*.
[39] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*.
[40] T.-Y. Hsiao, Y.-C. Chang, H.-H. Chou, and C.-T. Chiu, "Filter-based deep-compression with global average pooling for convolutional networks," *J. Syst. Architecture*, vol. 95, Feb. 2019, doi: 10.1016/j.sysarc.2019.02.008.
[41] N. Ketkar, "Introduction to keras," in *Deep Learning with Python*. Springer, 2017, pp. 97–111.
[42] M. B. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operat. Syst. Design Implement. OSDI*, 2016, pp. 265–283.

Binge Cui received the B.Sc., M.Sc., and Ph.D. degrees in computer science from Harbin Engineering University, Harbin, China, in 2000, 2003, and 2006, respectively.

In 2006, he joined the College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China. From 2010 to 2011, he was a Visiting Scholar with the Department of Information System, City University of Hong Kong. From 2012 to 2014, he was a Postdoctoral Researcher with the First Institute of Oceanography, State Oceanic Administration, China. He is currently a Professor. His research interests include hyperspectral image classification and remote sensing images understanding with deep learning.



Wei Jing received the B.M. degree in e-commerce from Shandong University of Science and Technology, in Qingdao, China, in 2019, where he is currently working toward the M.S. degree with the College of Computer Science and Engineering.

His current research interests include deep learning and remote sensing image processing.

Ling Huang received B.Sc. in computer science from Qufu Normal University, in Qufu, China, in 2001, and the M.Sc. degree in software and theory of computer from Northwestern Polytechnical University, Xian, China, in 2004.

In 2004, she joined the College of Computer Science and Engineering, Shandong University of Science and Technology. She is currently a Lecturer. Her research interest includes hyperspectral image classification.

Zhongrui Li received the B.Sc. degree in computer science and technology from Liaocheng University, in Liaocheng, China, in 2019. He is currently working toward the M.S. degree with the College of Computer Science and Engineering at Shandong University of Science and Technology, Qingdao, China.

His current research interests include deep learning and remote sensing image processing.

Yan Lu received the B.Sc. and M.Sc. degrees in computer science from Yanshan University, Qinhuangdao, China, in 1998 and 2000, respectively, and the Ph.D. degree in computer science from and Fudan University, Shanghai, China, in 2003.

From 2003 to 2005, she was a Postdoctoral Researcher in Computer Science and Technology with Harbin Institute of Technology, Harbin, China. In 2005, she joined the College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao, China. She is currently an Associate Professor. Her research interests include hyperspectral image classification and object detection.