DANet-SMIW: An Improved Model for Island Waterline Segmentation Based on DANet

Jiawei Xu, Jing Li, Xiaoyu Zhao, Kuifeng Luan, Congqin Yi, and Zhenhua Wang

***Abstract*—Segmentation of island waterline contributes to shoreline movement analyzing, environmental monitoring, and integrated coastal zone management. To achieve high efficiency and high accuracy of island waterline segmentation in remote sensing images, we proposed a model for island waterline segmentation based on DANet (DANet-SMIW). In DANet-SMIW model, different indexes (Normalized Difference Water Index and OTSU) were taken as new channels added to input dataset, which enhanced waterline's spectral information. DANet backbone network was improved by dense connection of DenseNet. Loss function, consisting of binary cross entropy loss and Dice loss, was used to resolve the sample imbalance problem, and then, rough results of island waterline segmentation were refined by boundary refinement module (BRM). In total, 2042 island images were taken as experiment dataset, which were cropped from Landsat-8 images and divided into 1634 images for training, 100 images for testing, and 308 images for validation, and DANet-SMIW model was compared against other models, including FCN-32s, DeepLabv3+, PSPNet, Dense-ASPP, PSANet, ICNet, DuNet, and PIDNet. Results demonstrated that DANet-SMIW model achieved the highest values with pixel accuracy and Mean Intersection over Union and possessed higher segmentation efficiency than most other models. Collectively, DANet-SMIW model was an integrated accurate and efficient model for island waterline segmentation in remote sensing images.**

***Index Terms*—DANet, deep learning, island waterline, semantic segmentation.**

I. Introduction

SHORELINE of island is the intersection of land with water surface, which is affected by waves, currents, tides, and winds, and thus, it is dynamic and difficult to segment accurately [1]. In contrast to shoreline, waterline of island is defined as boundary between water body and exposed tidal flat, and it is recognizable in remote sensing images [2]. In recent years, almost 70% of the world's beaches are experiencing coastal erosion due to sea level rise, coastal erosion and sedimentation, and overexploitation of resources [3]. Dynamic monitoring of the island waterline can provide a scientific solution. Therefore, high-efficiency and high-accuracy segmentation of island waterline contribute to a comprehensive analysis of shoreline movement, environmental monitoring, and integrated coastal zone management [4], [5].

Because the images captured by remote sensing technology have a large monitoring range, fast information acquisition, and low ground influence, it has been widely used for image segmentation research [6], [7]. Compared with the Suomi National Polar-orbiting Partnership and the Visible Infrared Imaging Radiometer Suite [8], Landsat-8 images are publicly available and have an order of magnitude improvement in resolution of 30 m. In this study, Landsat-8 images were used for island waterline segmentation.

Many methods had been applied to segment island waterline, such as edge detection method [9], index analysis method [10], thresholding method [11], wavelet transform method [12], etc. Among them, Niedermeier et al. [12] used wavelet and active contour methods to derive coastlines from SAR images by using. Liu and Jezek [13] presented a comprehensive approach based on a locally adaptive thresholding technique to effectively and accurately extract coastlines from satellite imagery. Cheng et al [14] proposed a graph cut-based supervised method to segment the sea and the land from natural-colored images. All these methods were based on manual annotation and crafted based on domain knowledge, which were time-consuming.

Deep learning models had been widely utilized for object segmentation due to their automatic learning capability and powerful predictive ability. For example, based on fully convolutional networks (FCN), several research works proposed end-to-end semantic segmentation models, which could take arbitrary input sizes and get corresponding output sizes [15], [16]. Based on encoder–decoder (ED), many research works fused multilevel semantic features and obtained abundant contextual semantic information, which could learn more classification information at the pixel level [17], [18]. Based on multiscale and pyramid (MSP), many studies captured global features and fused high and low-resolution features with minimal additional computational cost, which enabled a more comprehensive analysis of semantic and spatial information [19], [20], [21]. Based on DeepLab, many research works presented dilated convolution and atrous

Manuscript received 10 August 2023; revised 28 September 2023; accepted 28 October 2023. Date of publication 13 November 2023; date of current version 6 December 2023. This work was supported in part by Fujian Provincial Key Laboratory of Coast and Island Management Technology Study under Grant FJCIMTS 2023-04, in part by the Shanghai Ocean Bureau Research Project (Shanghai 2023-01) and Capacity Development for Local College Project under Grant 19050502100. *(Jiawei Xu, Jing Li, and Xiaoyu Zhao contributed equally to this work.) (Corresponding author: Zhenhua Wang.)*

Jiawei Xu, Xiaoyu Zhao, Kuifeng Luan, Congqin Yi, and Zhenhua Wang are with the Shanghai Ocean University, Shanghai 201306, China, and also with Fujian Provincial Key Laboratory of Coast and Island Management Technology Study, Xiamen 361013, China (e-mail: jwxu@shou.edu.cn; zhaoxiaoy99@gmail.com; kfluan@shou.edu.cn; cqyi@shou.edu.cn; zh-wang@shou.edu.cn).

Jing Li is with the Eye Institute and Department of Ophthalmology, Eye and ENT Hospital, Fudan University, Shanghai 200031, China (e-mail: 1jing.li@fdeent.org).

This article has supplementary downloadable material available at https://doi.org/10.1109/JSTARS.2023.3332427, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2023.3332427

TABLE I
ADVANTAGES AND DISADVANTAGES OF DIFFERENT DEEP LEARNING MODELS

Classification	Representative model	Advantages	Disadvantages
Based on FCN	FCN	Accepting images of any size, improving computational efficiency, combining depth information	Insufficient segmentation results, insensitivity to details, and insufficient consideration of the relationships between pixels
Based on ED	SegNent U-Net V-Net	End-to-end training, strong feature extraction ability, and high flexibility	Needing to adjust parameters
Based on MSP	PSPNet UperNet APCNet DenseASPP	Accepting input images of different scales and sizes, better capture the feature information of images	Needing to adjust input images of different scales and sizes, feature extraction, and the introduction of contextual information, requiring more computational resources and time consumption.
Based on DeepLab	DeepLab V1 DeepLab V2 DeepLab V3 DeepLab V3+	high accuracy, strong robustness	high computational complexity and needing to adjust parameters
Based on Attention	PSANet DANet CANet OCNet EMANet	Capturing local connections of elements in the input sequence, performing parallel calculations, and having low model complexity	Inability to capture word order information and easy weakening of long-distance information
Based on Transformer	SETR TransUNet SegFormer Segmenter	Strong feature capture capability and high computational efficiency	Poor parameter sharing features and low flexibility

spatial pyramid pooling operations, which could solve feature loss problems caused by downsampling, as well as the multi-scale problem [22], [23], [24], [25]. Based on attention, several studies applied the attention mechanism, which improved the ability of feature extraction [26], [27]. Based on Transformer, many research works utilized transformer methods for image processing and modeling from global relationships, which could capture relationships between global features and features and retain more spatial information [28], [29], [30]. Table I showed the advantages and disadvantages of different deep learning models. Thus, deep learning models have been applied to remote sensing images segmentation [31], [32]. There were still several challenges in using deep learning models for island waterline segmentation in remote sensing images, such as how to make full use of multispectral information of remote sensing images, how to improve segmentation accuracy of island waterline, especially for small-size islands segmentation, such as those with an area of less than 0.03 square kilometers, and how to solve sample imbalance problem where there were more negative samples than positive samples.

DANet was proposed by Fu et al. [33] to adaptively integrate local features and global dependencies by capturing global feature dependencies in the spatial and channel dimensions. The position attention module (PAM) and channel attention module (CAM) were used to filtrate key features. PAM extracted information from feature maps through a convolutional network, and then obtained attention weights for each position through matrix multiplication and softmax processing, which could rich contextual relationships into local features and capture new features of spatial long-range contextual information [see Fig. 1(a)]. CAM directly calculated the attention weights between channels, extracted the response strength of feature channels through a convolutional network, and then obtained the attention weights for each channel through matrix multiplication and softmax processing, which could dig into interdependencies between channels and captured long-range contextual information in the channel dimension [see Fig. 1(b)]. For island waterline segmentation, DANet was better suited to deal with multichannel remote sensing images. But the backbone of DANet had numerous parameters for feature extraction that affected the computational efficiency, and its repeated convolution operations caused feature loss, which limited the feature extraction capability of the model. DenseNet was proposed by Huang et al. [34], which improved the efficiency of information and gradient transmission in the network, reduced the loss of low-level features, and increased the utilization of features. Dilated convolution [35] was proposed for expanding the perceptual field and acquiring more contextual multiscale information without adding additional computational effort. Therefore, in the semantic segmentation task, dilated convolution often was used to compensate for the lost receptive field [36].

In this study, we proposed a segmentation model for island waterline segmentation on DANet (DANet-SMIW). In DANet-SMIW model, taking DANet as a basic semantic

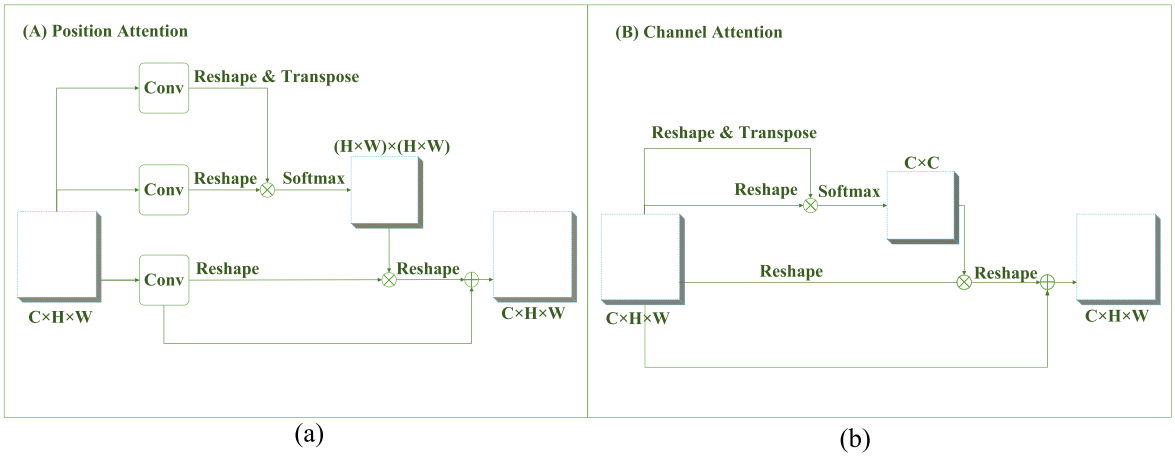


Fig. 1. Attention module. (a) PAM. (b) CAM.

TABLE II
CHARACTERISTICS OF LANDSAT-8 REMOTE SENSING IMAGES

Band	Wavelength（μm）	Spatial Resolution（m）	Application
Band1 Coastal	0.433–0.453	30	Coastal zone observation.
Band2 Blue	0.450–0.515	30	Water penetration and differentiation of soil vegetation.
Band3 Green	0.525–0.600	30	Distinguishing vegetation.
Band4 Red	0.630–0.680	30	Observing roads, bare soil, and vegetation species.
Band5 NIR	0.845–0.885	30	Estimating biomass and distinguishing wet soils.
Band6 SWIR1	1.560–1.660	30	Distinguishing roads, bare soil, and water; comparing different vegetation; distinguishing atmosphere, clouds, and fog
Band7 SWIR2	2.100–2.300	30	Distinguishing rocks, minerals, vegetation cover, and wet soils.
Band8 Pan	0.500–0.680	15	Enhancing resolution.
Band9 Cirrus	1.360–1.390	30	Including strong water vapor absorption features for cloud detection.

segmentation network, different indices [Normalized Difference Water Index (NDWI) and OTSU] were added as new channels to the input data, which enhanced spectral information of the island waterline and contributed DANet-SMIW model to segment out the island waterline. Improved DenseNet was taken as the backbone network of DANet-SMIW model, which could improve the feature extraction capability of DANet-SMIW model, especially for small-size islands. Binary cross entropy loss and Dice loss were used as the loss function of DANet-SMIW model, which solved sample imbalance problem and improved the accuracy of the island waterline segmentation. Lastly, the rough segmentation result was refined by the boundary refinement module (BRM), which corrected the segmentation boundary of the island waterline and reduced the probability of misclassification. The experimental results showed that the DANet-SMIW model could effectively segment out island waterline from remote sensing imagery.

II. Materials and Methodology

A. Study Area

The image dataset was acquired using Operational Land Imager (OLI) sensor in Landsat-8 and collected over Asia-Pacific from 1 January, 2015, to 1 November, 2020. This scene consists of nine bands with a spatial resolution, of 30 m. Fig. 2 showed an example of some islands located in Asia-Pacific.[1]

B. Landsat-8

The Landsat-8 satellite was launched on 11 February, 2013, and orbits the Earth in a Sun-synchronous near-polar orbit with an altitude of 705 km, an inclination angle of 98.2°, and a repetition period of 16 days, completing an Earth orbit every 99 min. The Landsat-8 satellite carries an OLI land imager with nine bands and an imaging width of 185×185 km (see Table II).

C. DANet-SMIW Structure

Fig. 3 showed the flowchart of DANet-SMIW. Taking DANet as basic semantic segmentation network, three modules were improved. In input module, values of NDWI and OTSU were considered as new channels added to input data. In segmentation module, DANet backbone network was replaced by improved DenseNet. In output module, the rough segmentation result was refined by BRM [37]. In addition, binary cross entropy loss and Dice loss were taken as the loss function.

[1][Online]. Available: https://glovis.usgs.gov/

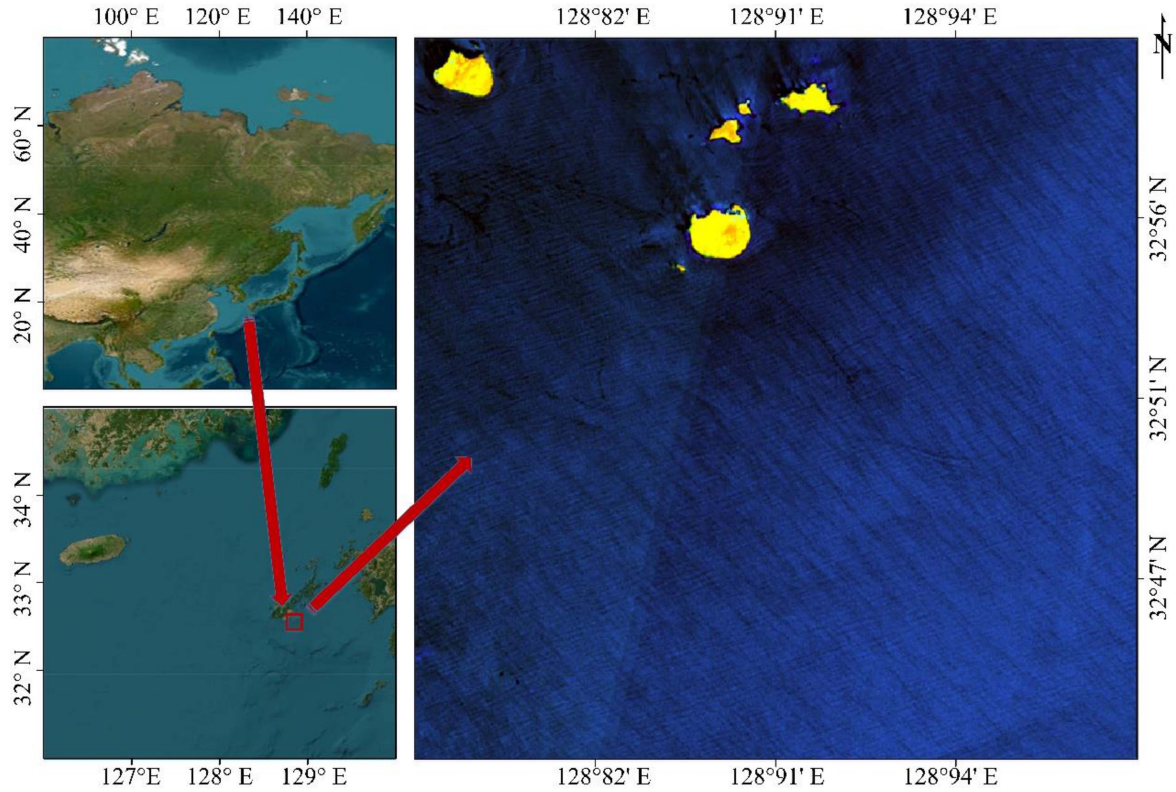


Fig. 2. Location of study area and example of data (bands 5, 6, and 4) from the study area.

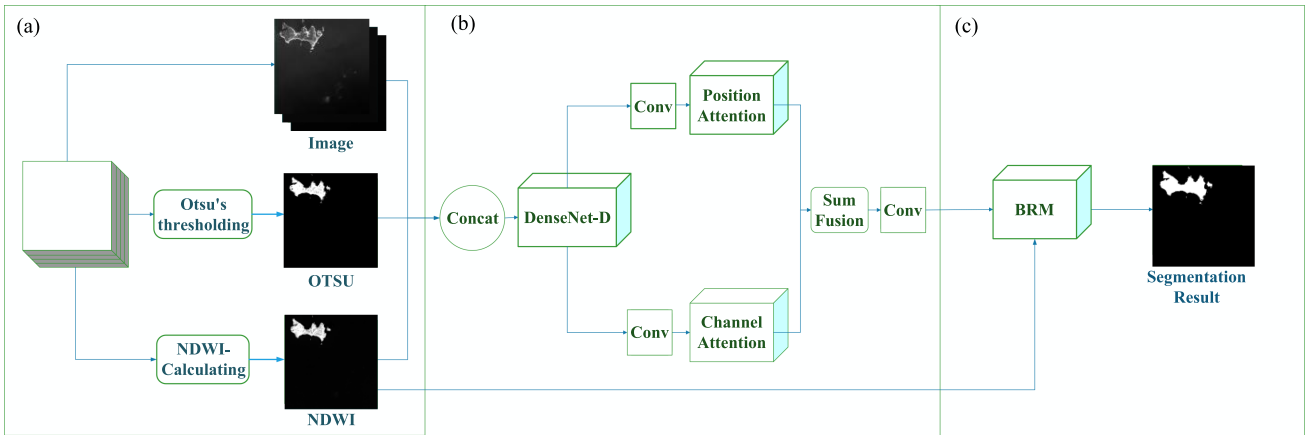


Fig. 3. Flowchart of DANet-SMIW. (a) Input module. (b) Segmentation module. (c) Output module. The concat represented that the different data were concatenated according to the channels, the DenseNet is the backbone network of the DANet-SMIW, the conv was the convolution operation, and the BRM is the boundary refinement module.

D. Taking Values of NDWI and OTSU as New Input Data Channels

DANet is designed for object segmentation in RGB images and has challenges for island waterline segmentation in remote sensing images, especially for multispectral features. Here, different indexes were used to enhance island waterline's spectral information, including NDWI and Otsu's thresholding method (OTSU).

NDWI is normalized difference processing with specific bands of remote sensing images to highlight the information about water bodies [38]. NDWI is calculated by

$$\mathrm{NDWI} = \frac{p\,(\mathrm{Green}) - p\,(\mathrm{NIR})}{p\,(\mathrm{Green}) + p\,(\mathrm{NIR})} \tag{1}$$

where $p(\mathrm{Green})$ represents the reflectivity of green band, and $p(\mathrm{NIR})$ the reflectivity of near-infrared band. NDWI values vary

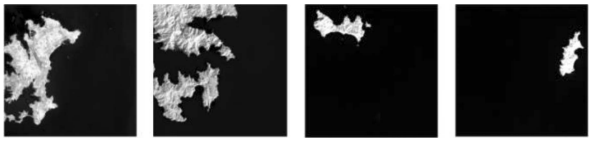


Fig. 4. Shape feature of islands extracted by NDWI.

Fig. 5. Rough segmentation result of island based on Otsu's thresholding.

in the range of -1 to 1, with open water features tending to have higher values greater than 0.

NDWI enhances the contrast between the water body and the background through spectral calculations. The value of NDWI was added to the input data of DANet-SMIW, which could enhance the shape feature of islands and inhibit part of background noise. Fig. 4 showed the shape feature of island segmented by NDWI. There were several fault segmentation results caused by shadow areas and omit segmentation for small-size islands. Thus, the rough result of island waterline segmented by Otsu was used as another input channel.

OTSU [39] separates the object from background by searching for the best threshold value. The optimal threshold value is found after performing a comprehensive search to maximize the variance between foreground and background. Variance between foreground and background g is defined as

$$g = w_0 * w_1 * (u_0 - u_1)^2 \tag{2}$$

where w_0 and $w_1$represent the percentage of foreground class pixels and background class pixels to image, respectively. And u_0 and u_1 are average gray values of foreground class and background class, respectively.

Fig. 5 showed rough results of islands waterline segmented by OTSU. OTSU could generate a rough segmentation result and provided identified foreground information with island waterline. However, the segmentation of OTSU cannot be directly used as the output result. Since OTSU needs to convert the original image into grayscale, which will lose part of the feature information, and while for multichannel images, the existence of a large amount of noise will also affect the segmentation results of OTSU.

In the input module of DANet-SMIW, both shape features of islands segmented by NDWI and rough results of island waterline segmented by OTSU were taken as new input channels. In addition, the shape feature of islands was also used as a feature vector to refine segmentation results in output module.

E. Improved DenseNet Backbone Network by Dilated Convolution

To retain more object details without adding an extra number of parameters, DenseNet was introduced to alleviate the vanishing gradient problem, enhance feature propagation, and substantially reduce the number of parameters. And in order to better extract features from remote sensing images, this backbone network was improved by dilated convolution (DenseNet-D). As seen in Fig. 6, 7×7 Conv-BN-ReLU-MaxPool operations were embedded after the input module. Then, dense blocks and transition blocks of DenseNet-121 network were chosen to segment semantic features of the island waterline, and dilated convolutions were employed in the last two dense blocks.

F. Refined Rough Segmentation Result By BRM

To enhance the accuracy of island waterline segmentation, a BRM was used to refine the rough segmentation result of island waterline in output module. Fig. 7 showed the structure of BRM. Taking the shape feature of island segmented by NDWI as guidance, the shape feature and output feature were combined. Feature vector was to supervise island waterline segmentation task, which is defined by generating average pooling operation, an attention vector by 1×1 convolution and sigmoid function, features were selected and refined by multiplication operation, and then the rough segmentation result of island waterline were refined by two 3×3 convolution-batch normalization ReLU.

G. Taking BCE Loss and Dice Loss as Loss Function of DANet-SMIW

Cross-entropy loss function of DANet is suitable for dealing with multilabel classification tasks. Island waterline segmentation is a binary segmentation and has a larger number of background samples. Loss function of DANet-SMIW was combined by two loss functions: BCE loss [40] and Dice loss [41].

BCE loss computes error for each pixel between prediction and ground truth, which is formulated by

$$\text{Loss}_{\text{BCE}} = -\frac{1}{|\Omega|}\sum_{i\in\Omega}\left[G_i \cdot \log P_i + (1 - G_i) \cdot \log\left(1 - P_i\right)\right] \tag{3}$$

where Ω represents the number of pixels in segmentation map, G_i is the label of the ith pixel, and P_i is the prediction mask of the ith pixel.

Dice loss is used to balance the positive and negative samples, which is defined by

$$\text{Loss}_{\text{Dice}} = 1 - \frac{2 \cdot G \cap P + \alpha}{|G| + |P| + \alpha} \tag{4}$$

where G and P are the label image and the prediction image, respectively. α can avoid that divisor that is 0 and overfitting phenomenon, which is optional. Here, the parameter α is set to 0.00001. DANet-SMIW uses BCE loss and Dice loss together

$$\text{Loss} = \text{Loss}_{\text{BCE}} + \text{Loss}_{\text{Dice}}. \tag{5}$$

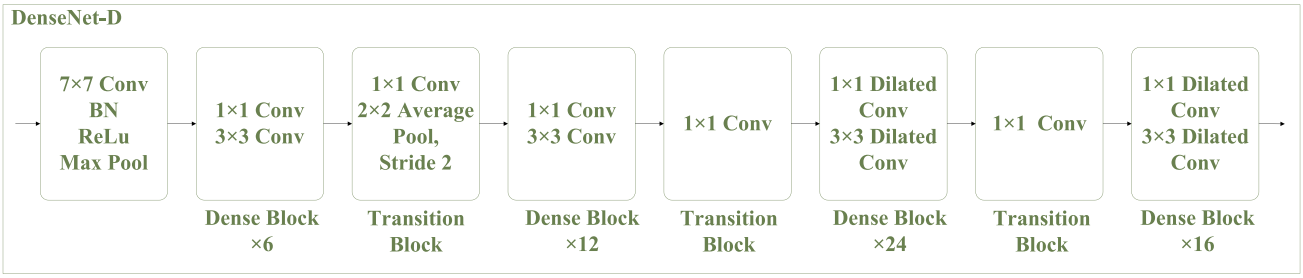


Fig. 6. Structure of improved DenseNet (DenseNet-D).

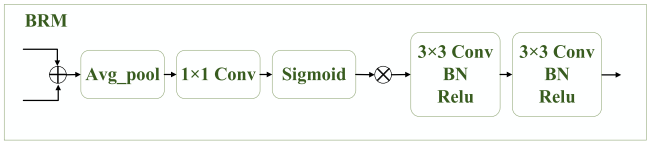


Fig. 7. Structure of BRM.

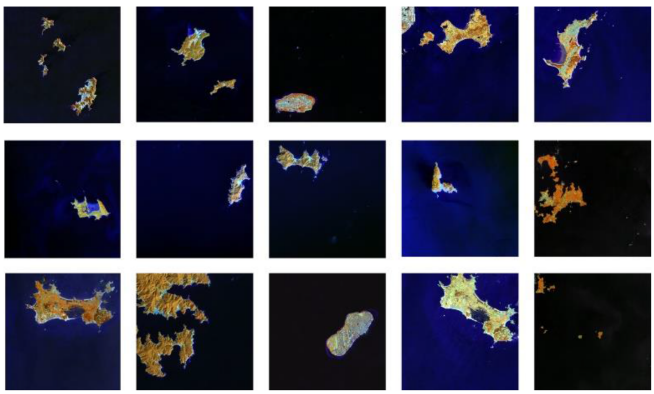


Fig. 8. Island images with bands 5, 6, and 4.

III. Experiments

A. Dataset

There were 2042 island images, which were randomly cropped at 480 × 480 pixels. In total, 2042 island images were split into 1634 images for training, 100 images for testing, and 308 images for validation. To ensure the validity of experimental results, training and testing images were independent. Fig. 8 showed the image of data cube with spectral bands 5, 6, and 4.

B. Experimental Setups

Here, deep learning models were trained for a minimum of 100 and a maximum of 300 epochs. And learning rate was set to 1e-4, and batch size was set to 2. All models are implemented using PyTorch 1.7.1+cu101 and executed on the Windows 10 platform with an NVIDIA Quadro RTX 3000 GPU.

C. Evaluation Metrics

Three evaluation metrics, including pixel accuracy (PA), mean Intersection over Union (mIoU), and frame per second (FPS), were calculated to estimate the performance of island waterline segmentation result [42].

PA represents the ratio of pixels properly classified, divided by the total number of pixels. For K classes, PA is defined as

$$\mathrm{PA} = \frac{\sum_{i=1}^{K+1} p_{ii}}{\sum_{i=1}^{K+1}\sum_{j=1}^{K+1} p_{ij}} \tag{6}$$

where $K+1$ classes include K foreground classes and the background class, and p_{ij} is the number of classes i predicted as class j. mIoU represents the average IoU over all classes, and IoU is the area of intersection between the predicted segmentation map and the label. mIoU is defined as

$$\mathrm{mIoU} = \frac{\sum_{i=1}^{K+1} \frac{|A_i \cap B_i|}{|A_i \cup B_i|}}{K+1} \tag{7}$$

where A and B denote the label and the predicted segmentation maps, respectively. FPS represents frames processed per second, which is used to evaluate the computation efficiency of models, and FPS is defined as

$$\mathrm{FPS} = \frac{1}{t} \tag{8}$$

where t represents the time of processing a picture.

IV. Results and Analysis

To evaluate the segmentation performance of DANet-SMIW model, two comparison experiments were designed. In ablation experiment, the loss function and each improved component were analyzed. In comparison experiment, the segmentation performance of DANet-SMIW model was compared with other models, including the FCN-32s [43], DeepLabv3+ [44], PSPNet [45], Dense-ASPP [46], PSANet [47], ICNet [48], DuNet [49], and PIDNet [50]. To ensure the fairness of the experiment, all models used the same training, validation, and testing sets, and applied the same data preprocessing and normalization techniques. In addition, for each model, the optimal parameter settings were used, and multiple experiments were conducted, with the average value as the final experimental result.

TABLE III
COMPARISON OF SEGMENTATION PERFORMANCE BETWEEN DANET-SMIW WITH DIFFERENT LOSS FUNCTION SETUPS

Models with different loss functions	PA	MIoU	FPS
DANet-SMIW with Cross Entropy Loss	97.38±0.06	91.62±0.53	14.15±0.28
DANet-SMIW with Binary Cross Entropy Loss	97.53±0.11	91.81±0.13	14.87±0.15
DANet-SMIW	97.86±0.30	92.31±1.13	14.51±0.31

TABLE IV
COMPARISON OF SEGMENTATION PERFORMANCE BETWEEN DANET-SMIW WITH DIFFERENT SETTING

RGB	Backbone		NDWI	OTSU	BRM	PA(%)	MIoU(%)	FPS(f/s)
	ResNet 50	DenseNet121						
√	√					95.88±0.08	89.46±0.28	8.61±0.07
√		√				97.86±0.30	92.31±1.13	14.51±0.31
√		√	√			98.46±0.08	94.37±0.17	14.32±0.46
√		√	√	√		98.63±0.09	95.12±0.14	13.21±0.30
√		√	√	√	√	99.08±0.13	96.36±0.15	12.52±0.05

TABLE V
COMPARISON OF SEGMENTATION PERFORMANCE AMONG DIFFERENT MODELS

Model	PA(%)	MIoU(%)	FPS(f/s)
FCN32s-VGG	92.89±0.34	78.53±0.39	19.57±0.40
DeepLabv3-Xception65	84.69±0.40	60.37±0.92	10.71±0.11
PSPNet-Resnet50	93.85±0.27	81.68±0.25	10.91±0.04
Denseaspp_Densenet121	95.39±0.22	84.76±0.22	16.73±0.11
PSANet-Resnet50	91.54±0.17	79.54±0.55	7.92±0.06
ICNet-Resnet50	93.46±0.09	83.44±0.45	22.43±0.17
DuNet-Resnet50	94.95±0.14	85.58±1.12	12.16±0.18
PIDNet	95.67±0.35	79.78±0.86	16.98±0.21
DANet-SMIW	99.08±0.13	96.36±0.15	12.52±0.05

A. *Ablation Experiment*

Table III showed the performance comparison of models with different loss functions, including DANet-SMIW with Cross Entropy Loss, DANet-SMIW with Binary Cross Entropy Loss, and our proposed DANet-SMIW. Compared with DANet-SMIW with Cross Entropy Loss, the evaluation metrics of DANet-SMIW, PA, mIoU, and FPS increased by 1.7%, 4.7%, and 0.4% respectively. Compared with DANet-SMIW with binary cross entropy loss, the evaluation metrics of DANet-SMIW, PA, and mIoU increased by 1.5% and 4.6%, respectively. And FPS of DANet-SMIW was slightly worse than DANet-SMIW with binary cross entropy loss. As a whole, combined with Cross Entropy Loss and Binary Cross Entropy Loss, DANet-SMIW could prominently enhance the accuracy and efficiency of island waterline segmentation.

Table IV showed the performance comparison between ablation studies. Taking the DenseNet as the backbone network, the evaluation metrics of DANet-SMIW, PA, mIoU, and FPS, reach 97.86%, 92.31%, and 14.51 f/s, respectively, which showed the applicability of DenseNet for feature extraction of remote sensing images. Augmented the input channel module by NDWI and OTSU, the evaluation metrics of DANet-SMIW, PA, mIoU, and FPS reach 98.63%, 95.12%, and 13.21f/s, respectively, which demonstrated that enhanced spectral features of the island waterline contributed to the DANet-SMIW model for segmentation. Used BRM to refine the segmentation result, the evaluation metrics of DANet-SMIW, PA, mIoU, and FPS reach 99.08%, 96.36%, and 12.52 f/s, respectively, which indicated the BRM could optimize the rough segmentation result. As a whole, the proposed model DANet-SMIW significantly improved the accuracy and efficiency of island waterline segmentation.

B. *Comparison Results*

The DANet-SWIM model was compared with FCN, DeepLabv3+, PSPNet, DenseASPP, PSANet, ICNet, DuNet, and PIDNet, on the island waterline dataset. Fig. 9 and Table V showed the comparison of segmentation performance among different models. The result demonstrated that DANet-SWIM

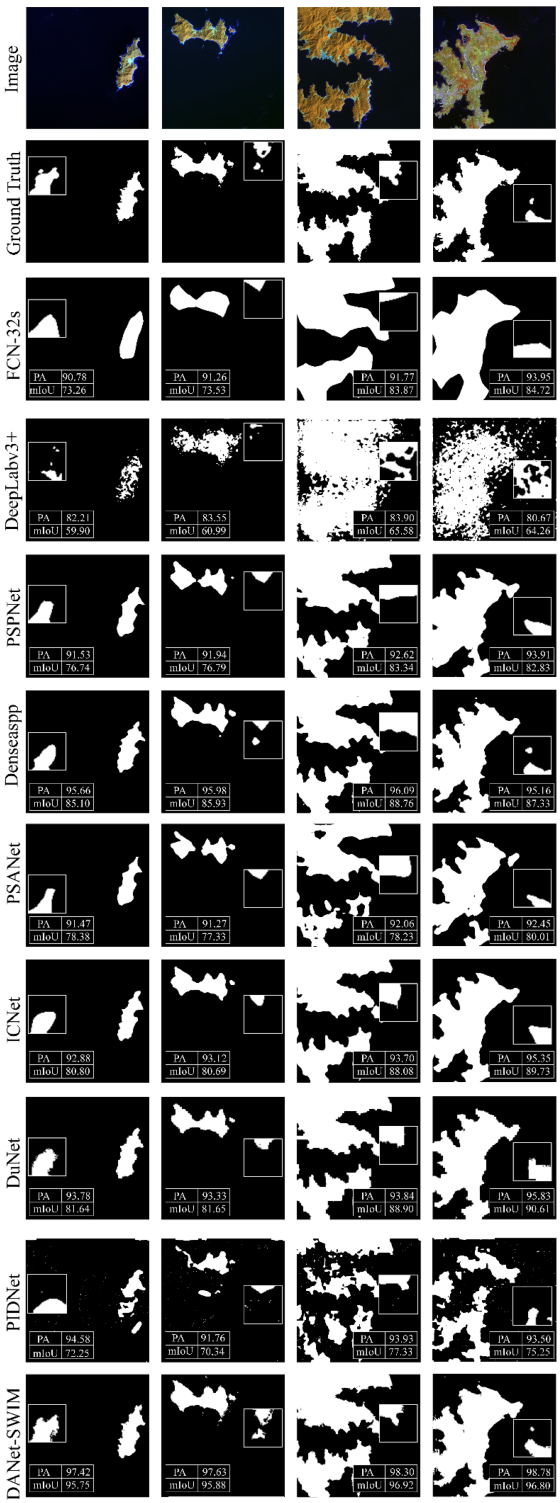


Fig. 9. Segmentation results of island waterline by different models.

model outperformed all other models and yielded the highest PA of 99.08±0.13 and the highest mIoU of 96.36±0.15. Compared with FCN, DeepLabv3+, PSPNet, DenseASPP, PSANet, ICNet, DuNet, and PIDNet, the PA value of DANet-SWIM was increased by 6.19%, 14.39%, 5.23%, 3.69%, 7.54%, 5.62%, 4.63%, and 3.41%, respectively, and the mIoU value of DANet-SWIM was increased by 17.83%, 35.99%, 14.68%, 14.60%, 16.82%, 12.92%, 10.78%, and 16.58%, respectively. The segmentation speed of DANet-SWIM was faster than that of the DeepLabv3+, PSPNet, PSANet, DuNet, and PIDNet, but was slightly slower than that of FCN-32s, Dense-ASPP, and ICNet.

In addition, the visualization results could be found in Fig. 9. Fig. 9 provided a more intuitive illustration that DANet-SMIW model had better segmentation performance. For example, the first and third columns indicated that only the segmentation results of the DANet-SWIM model were more consistent with the boundary contours of the labels. The second and fourth columns showed that only DenseASPP and DANet-SWIM models could segment the small-sized island waterlines. While the segmentation results of DANet-SWIM were closer to the labels.

In conclusion, DANet-SWIM showed powerful multispectral data analysis capabilities and effectively reduced false segmentation rate.

V. Discussion

The advent of era of Big Data promotes the development and application of remote sensing. And remote sensing images have been used in surveillance, security, traffic and environmental monitoring, and so on [51]. At the same time, with continuous innovation of deep learning technology, there are many excellent semantic segmentation models proposed.

Island waterline segmentation in remote sensing images is particularly valuable for marine monitoring and sea island management. Traditionally, manual monitoring is time-intensive and expensive [52]. In recent years, deep learning has made remarkable progress in the field of image segmentation. And some deep learning models were used to island waterlines extraction [53], [54]. However, severe challenges, such as the inability to fully utilize the multispectral information of remote sensing images, and widely distributed small-size islands limit the performance of deep learning models to a great extent. Therefore, we designed a deep learning model for island waterline segmentation in remote sensing images.

In the input module, the addition of NDWI and OTSU channels enabled DANet-SMIW model to obtain more information on the spectral features of island waterline. Compare with the model using only RGB data, the DANet-SMIW model presented a more accurate segmentation of island waterline, as shown in Table III. The additional NDWI and OTSU channels contributed to the full utilization of multispectral images to improve the accuracy of segmentation result.

In the segmentation module, the improved DenseNet could better extract object features. Benefitting from the feature extraction capability of the improved DenseNet, the DANet- SMIW model was able to segment out island waterline with higher accuracy and efficiency, as confirmed in Table III. In addition, it also enhanced the ability for small-size island waterline segmentation.

In the output module, the shape feature of island segmented by NDWI was taken as guidance, and the BRM was introduced to refine the rough segmentation result of island waterline. Through

the refinement, the DANet-SMIW model further provided better accuracy of island waterline segmentation.

In addition, the sample imbalance problem was addressed by two loss functions, as shown in Table II. Overall, the DANet-SMIW model achieved the best performance compared with other semantic segmentation models, as shown in Fig. 9 and Table V.

Nonetheless, there is still some work to be done in our study. The DANet-SMIW model shows great potential in improving the accuracy of segmentation results, but the efficiency is slightly slower than ICNet and DenseASPP. In future work, we will further improve DANet-SMIW model to enhance the accuracy and efficiency of island waterline. And combined with the tidal information, we will attempt to segment the coastline of island.

VI. Conclusion

In this study, a segmentation model DANet-SMIW was proposed for island waterlines segmentation in remote sensing images. To take advantage of multispectral of the remote sensing images, the input module is augmented by NDWI and OTSU. To overcome the omit segmentation phenomenon for small-size island, the segmentation module was improved by dense connection. To refine the island waterline, the output module is optimized by BRM. And then, taking binary cross entropy loss and Dice loss as the loss function, DANet-SMIW was improved to solve the problem of samples distribution imbalance. Compared with other segmentation models, including the FCN-32s, DeepLabv3+, PSPNet, Dense-ASPP, PSANet, ICNet, and DuNet, our proposed DANet-SMIW model could significantly enhance segmentation accuracy. And in the future, we plan to accomplish dynamic monitoring of the island waterlines.

References

[1] U. Natesan, A. Parthasarathy, R. Vishnunath, G. Kumar, and V. Ferrer, "Monitoring longterm shoreline changes along Tamil Nadu," *India Using Geospatial Techn.*, vol. 4, pp. 325–332, 2015.

[2] D. Mason, D. Hill, I. Davenport, R. Flather, and G. Robinson, "Improving inter-tidal digital elevation models constructed by the waterline technique," *ESA SP (Print)*, vol. 414, pp. 1079–1082, 1997.

[3] M. K. Ghosh, L. Kumar, and C. Roy, "Monitoring the coastline change of Hatiya Island in Bangladesh using remote sensing techniques," *ISPRS J. Photogrammetry Remote Sens.*, vol. 101, pp. 137–144, 2015.

[4] C. H. Fletcher et al., "National assessment of shoreline change: Historical shoreline change in the Hawaiian Islands," U.S. Geol. Survey, U.S. Dept. Interior, Washington, DC, USA, Tech. Rep. 2011-1051, , pp. 1051–1055, 2012.

[5] T. Kongeswaran and R. Karikalan, "Mapping of shoreline changes in between Devipattinam and Kilakkarai, Tamilnadu," *Southeast Coast India*, vol. 2, pp. 12–19, 2016.

[6] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," *Expert Syst. Appl.*, vol. 169, 2021, Art. no. 114417.

[7] Y. Wang, X. Jing, Y. Xu, L. Cui, Q. Zhang, and H. Li, "Geometry-guided semantic segmentation for post-earthquake buildings using optical remote sensing images," *Earthq. Eng. Struct. Dyn.*, vol. 52, pp. 3392–3413, 2023.

[8] Z. Wang et al., "Semantic segmentation and analysis on sensitive parameters of forest fire smoke using smoke-unet and Landsat-8 imagery," *Remote Sens.*, vol. 14, 2021, Art. no. 45.

[9] 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, pp. 937–958, 2004.

[10] T. Kuleli, A. Guneroglu, F. Karsli, and M. Dihkan, "Automatic detection of shoreline change on coastal Ramsar wetlands of Turkey," *Ocean Eng.*, vol. 38, pp. 1141–1149, 2011.

[11] S. K. Pal, A. Ghosh, and B. Shankar, "Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation," *Int. J. Remote Sens.*, vol. 21, pp. 2269–2300, 2000.

[12] A. Niedermeier, E. Romaneessen, and S. Lehner, "Detection of coastlines in SAR images using wavelet methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2270–2281, Sep. 2000.

[13] 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, pp. 937–958, 2004.

[14] D. Cheng, G. Meng, S. Xiang, and C. Pan, "Efficient sea–land segmentation using seeds learning and edge directed graph cut," *Neurocomputing*, vol. 207, pp. 36–47, 2016.

[15] G. Chen et al., "SDFCNv2: An improved FCN framework for remote sensing images semantic segmentation," *Remote Sens.*, vol. 13, 2021, Art. no. 4902.

[16] Y. Wu, Y. Qin, Y. Qian, F. Guo, Z. Wang, and L. Jia, "Hybrid deep learning architecture for rail surface segmentation and surface defect detection," *Comput.-Aided Civil Infrastructure Eng.*, vol. 37, pp. 227–244, 2022.

[17] M. Alam, J.-F. Wang, C. Guangpei, L. Yunrong, and Y. Chen, "Convolutional neural network for the semantic segmentation of remote sensing images," *Mobile Netw. Appl.*, vol. 26, pp. 200–215, 2021.

[18] X. He, Y. Zhou, J. Zhao, D. Zhang, R. Yao, and Y. Xue, "Swin transformer embedding UNet for remote sensing image semantic segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4408715.

[19] M. Xia, Y. Cui, Y. Zhang, Y. Xu, J. Liu, and Y. Xu, "DAU-Net: A novel water areas segmentation structure for remote sensing image," *Int. J. Remote Sens.*, vol. 42, pp. 2594–2621, 2021.

[20] B. Yu, L. Yang, and F. Chen, "Semantic segmentation for high spatial resolution remote sensing images based on convolution neural network and pyramid pooling module," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3252–3261, Sep. 2018.

[21] X. Gao et al., "An end-to-end neural network for road extraction from remote sensing imagery by multiple feature pyramid network," *IEEE Access*, vol. 6, pp. 39401–39414, 2018.

[22] N. Venugopal, "Automatic semantic segmentation with DeepLab dilated learning network for change detection in remote sensing images," *Neural Process. Lett.*, vol. 51, pp. 2355–2377, 2020.

[23] Y. Wang et al., "Geoinformation. Mask DeepLab: End-to-end image segmentation for change detection in high-resolution remote sensing images," *Int. J. Appl. Earth Observ. Geoinform.*, vol. 104, 2021, Art. no. 102582.

[24] S. Du, J. Xing, J. Li, S. Du, C. Zhang, and Y. Sun, "Open-pit mine extraction from very high-resolution remote sensing images using OM-DeepLab," *Natural Resour. Res.*, vol. 31, pp. 3173–3194, 2022.

[25] B. Quan, B. Liu, D. Fu, H. Chen, and X. Liu, "Improved Deeplabv3 for better road segmentation in remote sensing images," in *Proc. Int. Conf. Comput. Eng. Artif. Intell.*, 2021, pp. 331–334.

[26] B. Wu, Y. Gao, X. Xie, and Y. Yang, "Research and application of automatic recognition and extraction of water body in remote sensing images based on deep learning," in *Proc. Wirel. Technol., Intell. Netw. Technol., Smart Serv. Appl.: 4th Int. Conf. Wirel. Commun. Appl.*, 2022, pp. 167–173.

[27] S. Huang, W. Han, H. Chen, G. Li, and J. Tang, "Recognizing zucchinis intercropped with sunflowers in UAV visible images using an improved method based on OCRNet," *Remote Sens.*, vol. 13, 2021, Art. no. 2706.

[28] C. Zhang, W. Jiang, Y. Zhang, W. Wang, Q. Zhao, and C. Wang, "Transformer and CNN hybrid deep neural network for semantic segmentation of very-high-resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 4408820.

[29] D. Feng, Z. Zhang, and K. Yan, "A semantic segmentation method for remote sensing images based on the Swin transformer fusion Gabor filter," *IEEE Access*, vol. 10, pp. 77432–77451, 2022.

[30] L. Wang, R. Li, C. Duan, C. Zhang, X. Meng, and S. Fang, "A novel transformer based semantic segmentation scheme for fine-resolution remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Feb. 2022, Art. no. 6506105.

[31] L. Cui, X. Jing, Y. Wang, Y. Huan, Y. Xu, and Q. Zhang, "Improved swin transformer-based semantic segmentation of postearthquake dense buildings in urban areas using remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 369–385, Nov. 2023.

[32] Y. Wang, L. Cui, C. Zhang, W. Chen, Y. Xu, and Q. Zhang, "A two-stage seismic damage assessment method for small, dense, and imbalanced buildings in remote sensing images," *Remote Sens.*, vol. 14, 2022, Art. no. 1012.

[33] J. Fu et al., "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3141–3149.
[34] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.
[35] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*.
[36] Y. Zhang, J. Liu, and K. Huang, "Dilated hourglass networks for human pose estimation," in *Proc. Chin. Automat. Congr.*, 2018, pp. 2597–2602.
[37] Z. Zhao, C. Xia, C. Xie, and J. Li, "Complementary trilateral decoder for fast and accurate salient object detection," in *Proc. 29th ACM Int. Conf. Multimedia*, 2021, pp. 4967–4975.
[38] B. Gao, "NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space," *Remote Sens. Environ.*, vol. 58, pp. 257–266, 1996.
[39] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
[40] P.-T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Operations Res.*, vol. 134, pp. 19–67, 2005.
[41] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis.*, 2016, pp. 565–571.
[42] R. Nova, S. Nurmaini, R. U. Partan, and S. T. Putra, "Automated image segmentation for cardiac septal defects based on contour region with convolutional neural networks: A preliminary study," *Inform. Med. Unlocked*, vol. 24, 2021, Art. no. 100601.
[43] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.
[44] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.
[45] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6230–6239.
[46] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, "DenseASPP for semantic segmentation in street scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3684–3692.
[47] H. Zhao et al., "PsaNet: Point-wise spatial attention network for scene parsing," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 267–283.
[48] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, "ICNet for real-time semantic segmentation on high-resolution images," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 405–420.
[49] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A deformable network for retinal vessel segmentation," *Knowl.-Based Syst.*, vol. 178, pp. 149–162, 2019.
[50] T. J. Ma, "Remote sensing detection enhancement," *J. Big Data*, vol. 8, pp. 1–13, 2021.
[51] J. Xu, Z. Xiong, and S. P. Bhattacharyya, "PIDNet: A real-time semantic segmentation network inspired by PID controllers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 19529–19539.
[52] A. J. McLeay et al., "Deep convolutional neural networks with transfer learning for waterline detection in mussel farms," in *Proc. IEEE Symp. Ser. Comput. Intell.*, 2021, pp. 1–8.
[53] B. Cui, W. Jing, L. Huang, Z. Li, and Y. Lu, "SANet: A sea–land segmentation network via adaptive multiscale feature learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 116–126, Nov. 2021.
[54] E. Elkhateeb, H. Soliman, A. Atwan, M. Elmogy, K.-S. Kwak, and N. Mekky, "A novel coarse-to-Fine Sea-land segmentation technique based on superpixel fuzzy C-means clustering and modified Chan-Vese model," *IEEE Access*, vol. 9, pp. 53902–53919, 2021.

Jiawei Xu received the B.A. degree in aquaculture science from Shanghai Fisheries College, Shanghai, China, in 2007, and the M.S. degree in aquatic germplasm resources from Shanghai Ocean University, Shanghai, China, in 2010.

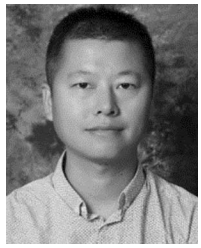
She is currently an Assistant Researcher with Shanghai Ocean University. Her current research interests include discipline analysis and higher education research.

Jing Li received the master's degree in computer science and technology from Shanghai Ocean University, Shanghai, China, in 2023.

Her research focuses on deep learning image processing.

Xiaoyu Zhao received the B.S. degree in spatial information and digital technology from Shanghai Ocean University, Shanghai, China, in 2021. She is currently working toward the M.S. degree in computer science and technology at Shanghai Ocean University, Shanghai, China.

Her research interests include assessment accuracy of spatial image and image processing with deep learning.



Kuifeng Luan received the Ph.D. degree in cartography and geographic information from Tongji University, Shanghai, China, in 2018.

He is currently an Associate Professor with the College of Marine Sciences, Shanghai Ocean University, Shanghai. His research interests include LiDAR remote sensing, and its applications in ocean and deep space exploration and geographic information.

Congqin Yi received the B.S. degree in applied electronic technology from the College of Electrical Engineering, Guizhou University, Guiyang, China, in 2000, the M.S. degree in computer science and technology from Guizhou University in 2003, and the Ph.D. degree in fishery economy management from Shanghai Ocean University, Shanghai, China, in 2013.

She is currently a lecturer with the College of Information Technology, Shanghai Ocean University. Her current research interests include data analysis, deep learning, and computer vision.

Zhenhua Wang received the B.S. degree in geographical science from the School of Geographical Sciences, Yantai Normal University, Yantai, China, in 2004, the M.S. degree in geochemistry from the Chinese Academy of Sciences, Beijing, China, in 2007, and the Ph.D. degree in cartography and geographic information engineering from Tongji University, Shanghai, China, in 2010.

She is currently a Professor with the College of Information Technology, Shanghai Ocean University, Shanghai. Her current research interests include spatial data analysis, data quality control, and deep learning.