Quadtree Decomposition-based Deep Learning Method for Multiscale Coastline Extraction with High-Resolution Remote Sensing Imagery

=Authors=

Shuting Sun[1], Lin Mu[2,3], Ruyi Feng[4], Yifu Chen[4], and Lizhe Wang[4]

[1]Collage of Marine Science and Technology, China University of Geosciences, Wuhan, P. R. China

[2]College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, P. R. China

[3]Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, China

[4]School of Computer Science, China University of Geosciences, Wuhan, P. R. China

Abstract The coastal zone is one of the most important features on the earth's surface; therefore, it is imperative to extract the coastline, a representative coastal zone feature with high quality. Previously, related methods mainly focus on edge and small-scale information. However, when processing large scale images, misclassification can occur because it's difficult to determine whether a local area belongs to land or sea. To solve this problem, in this study, a deep learning-based multiscale coastal line extraction algorithm is proposed, whose core is a scene classification-based multiscale coastal zone classifier to identify the coastal zones from low to high levels using quadtree decomposition. Compared

---

to the conventional method, the proposed method can obtain information from a large receptive field so as to identify land and sea precisely in high resolution imagery. The results indicate that the proposed method can effectively eliminate confusing features, and is of high calculation speed.

Plain Language Summary The coastal zone is one of the most important features on the earth's surface; therefore, it is imperative to extract the coastline, a representative coastal zone feature with high quality. Previous methods mainly focus on edge and small-scale information, which is able to produce results with accurate contours. However, when processing large scale images, misclassification can occur because it's difficult to determine whether a local area belongs to land or sea. To solve this problem, in this study, a deep learning-based multiscale coastal line extraction algorithm is proposed, whose core is a scene classification-based multiscale coastal zone classifier to identify the coastal zones from low to high levels using quadtree decomposition. Compared to the conventional method, the scene classification network can accurately identify land and sea owing to its larger receptive field. This advantage can be further strengthened by processing images with lower resolution and larger coverage using the quadtree decomposition structure. The results indicate that the proposed method can effectively eliminate confusing features, and the overall accuracy is improved by 5%. In addition, the screening process also significantly reduced the number of input samples for the segmentation network, which led to a higher calculation speed.

1. Introduction

The coastal zone is the interface between land and ocean. Because the majority of the world's population inhabits coastal zones, it is one of the most important features on the earth's surface; therefore, effective coastal zone management is essential to the human economy, society, and security. The coastline marks the transition between land and sea. It is the most representative feature of the coastal zone. Investigation of coastlines has attracted attention in coastal zone management.

Remote sensing techniques provide a new access to coastal zone investigations owing to their fast information acquisition speed, short data acquisition period, and mass information features. Because of these advantages, coastline extraction using high-resolution imagery has attracted scholarly attention and various coastline extraction methods have been developed.

Because the spectral characteristics of ground objects are obvious in early low-resolution remote sensing images, initial coastline extraction methods are mostly based on thresholding, index, operator, and other simple image processing methods. For instance, using a histogram thresholding method on band five of TM or ETM+ imagery(Bagli & Soille, 2003), an NDWI was imported to extract the coastline based on the reflectance of water and other land covers on each band of the TM or ETM+ imagery(Grigio et al., 2005; Ouma & Tateishi, 2006). Edge detection operators include extracting coastlines by integrating canny edge detection(Lin et al., 2003), integrating a canny edge detector with two masking steps(Heene & Gautama, 2000), and integrating locally adaptive thresholding methods(H. Liu & Jezek, 2004). These pixel-based methods only employ spectral information of individual pixels and ignore the relationship between neighboring pixels. They are simple and stable, which ensures their application in the production of long-term low-resolution coastline products. However, such simple methods do not perform well in complex scenarios such as images with high-resolution images or large coverage areas owing to the confusion caused by synonyms spectrum.

To solve the problem of inaccurate boundaries (the direct problem synonyms brought by the spectrum), subsequent studies focus on small-scale features using texture analysis and object-oriented classification models. Texture analysis uses GLCM, LBP, frequency domain transform, and other methods to calculate the texture features of each pixel(Bo et al., 2001; Wei & He, 2016; L. Li et al., 2021). Pixels with the same texture features are considered as homogeneous and merged into one region. Object-oriented classification methods divide coastal zone images into objects through segmentation, such as watershed or meanshift(Yang & Liu, 2005; Hollenhorst et al., 2007; Jarabo-Amores et al., 2010) and classify these objects as land and sea. Because the features are extended from pixel to neighborhood, these methods can extract accurate outlines on local high-resolution images; however, the utilization of spatial information is still limited and largely depends on manually selected parameters and artificial rules. When the study area becomes larger, a number of confusing objects will occur, and more complex artificial rules will be required. Such disadvantages limit the generality and robustness of these methods.

In recent studies, to address the problem of confusing objects in large coverage images, scholars have used deep learning methods to further improve the ability of image feature mining. Because the purpose of the coastline extraction is to extract complete coastline contours and semantic segmentation, an end-to-end network is the most suitable and widely adopted method. Similar to image segmentation methods, semantic segmentation networks can also extract accurate contours from high-resolution images. For example, FCN, a typical semantic segmentation model, is applied to coastline extraction (Pashaei & Starek, 2019). To enhance the effect of edge extraction, SeNet (Cheng et al., 2016) proposed multi-task loss by adding an edge loss function to the conventional loss function as the measurement of edge detection.

Considering the confusing objects problem, using series of convolution and deconvolution operations, a semantic segmentation network can obtain image features at multiple scales to analyze each pixel more comprehensively and accurately. Based on this, improvements in the network structure, loss function, and post-processing process are proposed to amplify the perception of features on a certain scale. For example, to reduce the misclassification, the OAE-V3 model is proposed. It integrates the RESNET structure into a conventional Deeplab V3 network (Sui et al., 2020). DeepUnet (R. Li et al., 2018) adds a downscaling process to the network and widens the input data to increase the receptive field of the deep learning network. The pixels of maritime imagery are divided into three classes: sea, land, and ships. The input is divided into fine- and coarse-scale paths and are joined together to extract the features of different scales (Lin et al., 2017).

Despite many efforts, because these methods are based on a single network and their receptive field cannot be expanded infinitely owing to the feasibility of increasing the network size, misclassification still occurs. For instance, high-resolution images require a very large-scale to distinguish between coastal and large reservoirs.

Currently, pre- and post-processing are added to compensate for this deficiency. For example, Res-UNet CRF (Chu et al., 2019) uses a conditional random field (CRF) model and morphological operation as post-processing to eliminate confusing features. A large amount of calculation is required for semantic segmentation on high-resolution remote sensing images, and confusing features are mainly located inland. This indicates that they have obvious distribution characteristics. Therefore, compared to the post-processing, eliminating confusing features in advance through pre-processing is more sophisticated. Considering the preprocessing methods, images of several specific scales were selected, and small-scale images were used to optimize the contour. Large-scale images were used to determine the feature category. For example, the original image is decomposed into small tiles for the initial segmentation and processing of the final segment based on the spatial key of each tile (Wang et al., 2020). This can extracting raw outlines of lakes in a large space through Landsat images and adjust the details of these outlines through GF-1 images (F. Chen et al., 2017).

In this study, we would further improve the preprocessing and propose a multiscale coastline extraction method. As a state-of-the-art machine learning method, deep learning plays an important role in our model. Two branches, scene classification and semantic segmentation, are related to our study.

Scene image classification refers to the task of grouping a whole picture into an exact semantic category. Considering previous studies, scholars proposed deep convolutional one-way networks such as Alexnet (Krizhevsky et al., 2012), ResNet (He et al., 2016), and VGG16 (Simonyan & Zisserman, 2014). These groundbreaking studies also proposed various methods, such as downscaling, dropout, and transformation on input labels. Subsequent studies attempted to design more elaborate structures rather than simply stacking convolutional layers deeper and deeper. The most influential is the inception family (Szegedy et al., 2015). It creates an inception module that performs three different convolutional processes on an input, and the outputs are concatenated and sent to the next inception module. Considering the latest inception networks (Inception v4 and Inception-ResNet v2)(Szegedy et al., 2017), improvements such as kernel size and replacement pooling with residual connections have been made. The goal of the assignment is to classify images into coastal and non-coastal areas, which is a simple binary scenario. In this study, MobileNet (Howard et al., 2019), a typical lightweight network as a coastal zone classifier, is introduced. It uses depthwise separable convolution and lightweight depthwise convolutions to filter features in the intermediate expansion layer. Particularly, it requires a small amount of calculation and is easy to train.

Semantic segmentation aims to link each pixel in an image to a class label. Therefore, it replaces the fully connected layer used for classification in a scene recognition network with a convolution layer and forms a fully convolutional network, such as FCN (Long et al., 2015), the landmark of image segmentation. Because pooling will irreversibly lose detailed information, it creates a deconvolution and skips the connection structure to recover detailed features. However, the loss of detail caused by repeated pooling operations cannot be perfectly repaired by such a simple method, and the output of the FCN is a coarse segmentation map.

There are two kinds of studies addressing this problem. One is to perform convolution on an input with different kernel sizes instead of repeated pooling processes, such as PSPnet (Zhao et al., 2017)(Yu et al., 2018) and Deeplab family (L.-C. Chen et al., 2014)(L.-C. Chen et al., 2017)(L.-C. Chen et al., 2018). These methods use atrous convolution, which effectively enlarges the field of view through a single spectral kernel. To better incorporate multiscale information, the subsequent version adopts a fully connected CRF and SPP. The other is to improve the deconvolution process, and the representative model of this type is the Unet(Ronneberger et al., 2015), which is a simple and clear end-to-end convolutional structure. The Unet consists of two parts: a downsampling part with the same effect as pooling, which is composed of two identical convolutions and a rectified linear unit (ReLU), and an upsampling part, which uses the deconvolution layers to increase the resolution of the output. To preserve better details, deconvolution and skip connection processes were added after each downscaling process. Recently, its improved version, Unet++ (Zhou et al., 2018), has been proposed. This redesigns the skip connections to aggregate the features of varying semantic scales at the decoder sub-networks. In this study, to segment the coast image more accurately, Unet++, one of the forefront methods, is used as the final coastline extraction model.

2. Overview of the Methodology

This study aims to solve the problem of confusing features by further strengthening the ability of large-scale spatial information mining. Compared to semantic segmentation (a pixel-to-pixel network with a limited receptive field) the scene classification network extracts the information of the entire image as a coastal zone identifier. Because inland and coastal areas have different suitable scales for interpretation, a multiscale dataset is generated. Considering the fractal dimension characteristics and self-similarity of the coastline shape in the research area, a quadtree decomposition framework is proposed for multiscale image discrimination. Through this process, a number of inland and offshore areas can be excluded at a very large-scale, and the coastal zones can be delimited accurately and sharply.

Specifically, considering our method, the tile map service (TMS) is utilized to reconstruct the input data to obtain a multiscale dataset based on a quadtree structure. Subsequently, a multiscale coastal zone identifier with a quadtree as the skeleton and scene classification network as a classifier is constructed to eliminate non-coastal zone tile images level by level. Considering this preprocessing, only tile images at the highest level and precisely in the coastal zone were retained for subsequent semantic segmentation.

Process flow of the proposed method is as follow: we first generate a multiscale dataset ($Cz[w,h] = [C_{z1}[w,h], C_{z2}[w,h]...C_{zn}[w,h]]$) from the original image (C), where z is the level dimension according to the data organization mode of the tile map service (TMS). The width of the labels of each level is the same, and the resolution of the previous level is twice that of the subsequent level. Second, labels located in the coastal zone at the original resolution are screened out through the coastal zone recognition network proposed in this study, and we obtain a sub-dataset($c_{zn}[w,h] \subset C_{zn}[w,h]$). Finally, we process semantic segmentation on $c_{zn}[w,h]$ or $C_{zn}[w,h]$ and obtain the result of the coastline extraction ($W[w,h]$).

3. METHODOLOGY

3.1. Multiscale Dataset Generation Based on TMS

Conventional remote sensing image semantic segmentation tasks mainly divide an original image into series of fixed-size samples with one or few resolutions. Considering coastline extraction, suitable scales for different relevant tasks vary greatly. For example, extracting accurate contours requires small-scale information, whereas identifying land and sea positions of an area requires larger-scale information. Considering this scenario, a dataset with multiple resolutions is required. In addition, in previous multiscale datasets, samples of different scales are separated from each other, and indexes between the samples of different scales are needed.

To achieve this, the tile map service (TMS) (*TMS Rule @ONLINE*, n.d.) was introduced. It is a protocol for serving maps as tiles (i.e., splitting map up into a pyramid of images at multiple zoom levels). Each output block is a fixed-size image encoded according to the spatial position.

The division rules are shown in (1). A pixel located at (lon, lat) is divided into a file *z_x_y.png* at level z.

$$\begin{cases} x = 2^{z-1} \cdot (\frac{lon}{180} + 1) \\ y = 2^{z-1} \cdot (1 - \frac{ln[tan(lat \cdot \pi/180) + sec(lat \cdot \pi/180)]}{\pi}) \end{cases} \tag{1}$$

Regarding this data format, images of different scales are assimilated into series of tile images of the same size to classify or segment them directly through a single deep learning network. In addition, the quadtree index can be constructed using the tile number. This provides convenience for the quadtree decomposition-based coastal zone recognition method.

3.2. Quadtree Decomposition-based Index

Considering the previous section, a multiscale dataset based on TMS was created, and each scale produced a corresponding result. In this section, it is necessary to find an appropriate rule to integrate these results. Previous studies disassembled the task into multiple parts. Each part of the task had a corresponding scale and finally merged the results through simple rules (L. Liu et al., 2018). However, because our task is difficult to disassemble and there are huge scales differences within the dataset, improvements in the utilization of this multiscale information are needed.

The study area used here is the coast of Shenzhen, a metropolis located in the South China Sea, a marginal sea with the Eurasian continent as the boundary, where remote sensing image is composed of two large land covers, water body and land. The water body is completely and evenly distributed at the edges of the image. According to (Xiaohua et al., 2004), the fractal dimension of the Chinese mainland coastline is approximately 1.16. Considering the spatial distribution, such a region is of a certain degree of self-similarity; therefore, it is suitable to use an adaptive way to construct an index to make each area find its own suitable interpretation scale. Consequently, processing the quadtree decomposition (Shusterman & Feder, 1994) on this type of image is unlikely to cause misclassification or over-segmentation, and it can achieve the effect of a compressed perception. We propose a quadtree decomposition framework to perform an accurate and efficient screening of coastal zone images.

The quadtree decomposition divides the geographic space recursively into different levels of quadtree structures. If a block is determined to be decomposed, it is subdivided into four blocks (2 $\times$ 2), and the criterion is applied again to the subdivided blocks. This process is repeated iteratively until all the blocks are determined. The quadtree decomposition process is shown in Fig.S1, where P_{coast} is the image feature, L is the TMS level, and T_{coast} and T_L correspond to the thresholds of P_{coast} and L, respectively. Because quadtree decomposition subdivides an image into blocks that are more homogeneous than the image itself through the discriminator, the core of the method is the discriminator. The original discriminator is usually composed of simple grayscale statistics or texture information, such as a grayscale threshold or standard deviation. In this study, to identify the complex multiscale coastal zone tile image, a scene classification network was used as the discriminator.

3.3. Multiscale Coastal Zone Recognition

Considering the previous section, quadtree decomposition was identified as an index for multiscale coastal zone identification for the integrity of land, and the sea distribution of the study area makes it less likely to be misclassified. In this section, the scene classification network and current frontier classifier are utilized as multiscale coastal zone recognition algorithms known as the discriminator of quadtree decomposition.

As a deep learning network, the input data size of the scene classification network is fixed. This indicates that the coastal images with very different resolutions are processed through one network. The consistency of the coastal zone images at different scales was analyzed using a gray histogram as shown in Fig.S2. The results indicate that waterbodies of different scales are of spectral and texture consistency for the histogram distribution of each band. Comparatively, there are obvious differences between waterbody and land area.

To quantify the proportion of seawater on different levels of tile images and to analyze the impact of freshwater (the most confusing feature at each level), the proportion of seawater and waterbodies of tile images in the corresponding coastal zones of different scales are calculated as shown in Fig.S3. Among them, the proportion of seawater is extracted from the electronic navigational chart (*Electronic Navigational Chart @ONLINE*, n.d.), and that of waterbody is extracted from Amap (*Amap @ONLINE*, n.d.). It can be observed that there is no obvious change in the proportion of the seawater area in the tile images of the coastal zones at different levels. In addition, at a lower level, the proportion of seawater was similar to that of the waterbody. In contrast, at higher levels, freshwater occupies a certain area in the tile images; therefore, the proportion of waterbodies is larger than that of seawater. Consequently, the identification of coastal zones should be more accurate at large-scales, and our quadtree decomposition method is an iterative process from large-scale to small scale. Two scene recognition networks were adopted as multiscale coastal zone discriminators. Because the application scenario is a simple binary classification, MobileNet V3 has been adopted. In addition, to verify the accuracy, the latest network of the Inception family, Inception Res V2, was adopted. The labels of the coastal area used in this study were generated based on a slight modification of the electronic navigational chart. Therefore, they are essentially semantic segmentation samples. To convert it into a fuzzy classification label, the coastal zone index P_{coast} is proposed as (2), where $Area$ and $Area_{sea}$ are the area of the tile image and waterbody on the tile image, respectively.

$$P_{coast} = \sqrt{min(\frac{Area_{sea}}{Area}, 1 - \frac{Area_{sea}}{Area}) \times 2} \tag{2}$$

Considering the discriminator, the threshold discrimination method was used as $P_{coast} > T_{coast}$. Because confusing features and imperfect labels have a greater imp/act on large scale targets, the threshold is slightly larger at the large-scales. Considering the experiment, T_{coast} was taken based on $0.5 - 0.3 \times (z_{max} - z)/(z_{max} - z_{min})$, where z, z_{min}, and z_{max} are the current, minimum, and maximum levels, respectively.

3.4. Coastline Extraction through Semantic Segmentation

After screening using the multiscale coastal zone identifier proposed in this study, high-resolution samples in a small area around the coastal zone were obtained. Finally, the filtered samples, rather than the entire image, were used to train the semantic segmentation network for accurate coastline contour extraction. In our study, a nested U-Net (a cutting-edge method in the field of semantic segmentation) was adopted.

Considering the experiment, the coastline broke when some tile images were missing, or the local coastline was too close to the TMS splitting line. To make the result more complete, the results at different levels are output, and the missing parts in the high-level results are covered with the upper-level results.

4. Experiments and Analyses

4.1. Study Area

The research area is ShenZhen, which covers a total area of 1996.85 km^2. It is an important economic zone in China and is located in the Pearl River Delta, north of Hong Kong. Because of its large population, high economic status, and the impact of land reclamation on the natural environment during urbanization, it is important to investigate its coastal zone.

Shenzhen is a cosmopolitan city with a population of tens of millions. Therefore, a number of sand and mud deposition coasts have been developed as bathing beaches. Nevertheless, it is also an important international port, and the container handling capacity of Shenzhen Port ranks fourth in the world. Therefore, there are several artificial coasts. In addition, the Shenzhen area has national mangrove nature reserves. This is because of the economic status and diverse types of coastlines in Shenzhen.

4.2. Experimental Data

The input data consist of two parts. One is the satellite imagery of Shenzhen collected from Google Map imagery (*Google Map @ONLINE*, n.d.), and the other is the labels of the coastal area, generated from a slightly modified electronic navigational chart.

To satisfy the requirements of the quadtree decomposition, we directly adopted the data organization form of TMS and collected TMS images of ShenZhen from level 11 (resolution, 64 m) to level 17 (resolution, 1 m). Both the input and output are file lists of the tile images. The semantic segment was used to extract the coastline.

4.3. Accuracy Assessment

The main purpose of our study is to solve the misclassification phenomenon caused by the inability to identify coastal locations in large-scale and high-resolution images. Therefore, we adopted pixel accuracy (PA), a commonly used deep learning accuracy index, to evaluate the overall accuracy.

In addition, to evaluate the extraction effect of the detailed coastline contour, we introduced line matching(LM), defined as (3), to reflect the error size in the local normal direction.

$$LM = \frac{Area[Sea_{real} \triangle Sea_{predict}]}{L_{real}} \tag{3}$$

It represents the ratio of the intersection area of the real coastline and the extraction result to the length of the coastline, where Sea_{real} refers to the real seawater (or land) area, $Sea_{predict}$ refers to the related extraction result, $\triangle$ refers to the difference set, and L_{real} is the length of the real coastline.

4.4. Experimental Results

The presentation and evaluation of the experimental results consist of the following three parts: evaluation of quadtree decomposition, evaluation of the overall accuracy and local area, and time consumption assessment.

As shown in Fig.S4, the extraction result of the Unet++ directly on the original resolution image is often unable to identify confusing features such as lakes, ponds, and ships. In contrast, the proposed method can effectively filter ships in the ocean and most confusing features on land, including the Tiegang Reservoir (the largest freshwater target in the study area, covering an area of 64 km^2). The pixel accuracy of each method is correspondingly 87.9%(Unet++), 92.6%(Quadtree only) and 93.5%(Our Method). Considering the proposed method, the training dataset of semantic segmentation was also screened using the quadtree decomposition. The accuracy of this method was also the highest. In addition to directly using Unet++, an experiment that only adopted the quadtree decomposition as a coastline identifier and used all labels in the training semantic segmentation network served as the control. The results show that quadtree decomposition improves the accuracy by approximately 5%. In addition, according to the split line of the quadtree decomposition, more than 90% of coastal zone blocks at level 17 are correctly identified in both the mobileNet and inception. Nevertheless, 95.1% of the result area or 20.6% tile images are confirmed at level 15 or above. This indicates that the best scale for coastal zone identification is at least four times of the original resolution.

As shown in Fig.S5, the red, green, and blue lines represent the ground truth, Unet++ (control group), and proposed method, respectively. In the local region, the proposed method exhibits a better performance. Owing to the confusing features, such as freshwater or some buildings, they are very similar to seawater areas, considering the features. During the training, they are considered as land. This will mislead the network and lead to coarse segmentation results. In contrast, many confusing features are excluded using the quadtree decomposition; therefore more accurate contours can be extracted. It can be observed from the results that the LM of the method in this study has an improvement of 0.5, compared to the direct use of semantic segmentation.

The operating environment includes one Rt x 2080 ti graphics card, 64 GB memory, and an NVME SSD. Extraction directly through semantic segmentation will process 53728 effective tile images and take more than 15 min. The proposed method only takes approximately 1.5 min, which is ten times faster than the direct use of the semantic segmentation. The running time of the scene classification network is 44 s, processing 4794 tile images. The semantic segmentation process only takes 60 s because it only processes 3536 tile images as shown in Table S1.

5. Conclusion

Coastline extraction through remote sensing images is an important task in coastal zone investigations. When processing large scale and high-resolution coastal zone images, because the existing methods only extract the coastline contour based on the limited receptive field, misclassification has occurred. To solve this problem, a multiscale coastline extraction method is proposed based on sense classification network and quadtree decomposition.

The core of this method is to use a multiscale coastal zone discriminator for preprocessing. It embeds the scene classification network into the quadtree decomposition framework to realize the coastal zone region recognition level by level. This pre-processing can further enhance the receptive field to achieve high-accuracy coastal zone identification and effectively eliminate easily confused objects. Moreover, the process of eliminating confusing objects optimizes the training set, improving the subsequent semantic segmentation network. In addition, because quadtree decomposition is a top-down process, many areas have been excluded on a large-scale, which has greatly reduced the amount of calculation and time consumption.

The proposed method was tested in Shenzhen, China. The results indicate that the pixel accuracy has been improved by 5%. In addition, the number of tile images to be processed is greatly reduced to the corresponding time consumption.

References

Amap @ONLINE. (n.d.). Retrieved from `https://amap.com`

Bagli, S., & Soille, P. (2003). Morphological automatic extraction of pan-european coastline from landsat etm+ images. In *International symposium on gis and computer cartography for coastal zone management* (pp. 256–269).

Bo, G., Delleplane, S., & De Laurentiis, R. (2001). Coastline extraction in remotely sensed images by means of texture features analysis. In *Igarss 2001. scanning the present and resolving the future. proceedings. ieee 2001 international geoscience and remote sensing symposium (cat. no. 01ch37217)* (Vol. 3, pp. 1493–1495).

Chen, F., Zhang, M., Tian, B., & Li, Z. (2017). Extraction of glacial lake outlines in tibet plateau using landsat 8 imagery and google earth engine. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(9), 4002–4009.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2014). Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*.

Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the european conference on computer vision (eccv)* (pp. 801–818).

Cheng, D., Meng, G., Cheng, G., & Pan, C. (2016). Senet: Structured edge network for sea–land segmentation. *IEEE Geoscience and Remote Sensing Letters*, *14*(2), 247–251.

Chu, Z., Tian, T., Feng, R., & Wang, L. (2019). Sea-land segmentation with res-unet and fully connected crf. In *Igarss 2019-2019 ieee international geoscience and remote sensing symposium* (pp. 3840–3843).

Electronic navigational chart @ONLINE. (n.d.). Retrieved from `https://nauticalcharts.noaa.gov/charts/noaa-enc.html`

Google map @ONLINE. (n.d.). Retrieved from `http://map.google.com`

Grigio, A. M., Amaro, V. E., Vital, H., & Diodato, M. A. (2005). A method for coastline evolution analysis using gis and remote sensing—a case study from the guamaré city, northeast brazil. *Journal of Coastal Research*, 412–421.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770–778).

Heene, G., & Gautama, S. (2000). Optimisation of a coastline extraction algorithm for object-oriented matching of multisensor satellite imagery. In *Igarss 2000. ieee 2000 international geoscience and remote sensing symposium. taking the pulse of the planet: The role of remote sensing in managing the environment. proceedings (cat. no. 00ch37120)* (Vol. 6, pp. 2632–2634).

Hollenhorst, T. P., Brown, T. N., Johnson, L. B., Ciborowski, J. J., & Host, G. E. (2007). Methods for generating multi-scale watershed delineations for indicator development in great lake coastal ecosystems. *Journal of Great Lakes Research*, *33*, 13–26.

Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., ... others (2019). Searching for mobilenetv3. In *Proceedings of the ieee/cvf international conference on* *computer vision* (pp. 1314–1324).

Jarabo-Amores, P., Rosa-Zurera, M., de la Mata-Moya, D., Vicen-Bueno, R., & Maldonado-Bascon, S. (2010). Spatial-range mean-shift filtering and segmentation applied to sar images. *IEEE Transactions on Instrumentation and Measurement*, *60*(2), 584–597.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*, 1097–1105.

Li, L., Su, H., Du, Q., & Wu, T. (2021). A novel surface water index using local background information for long term and large-scale landsat images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *172*, 59–78.

Li, R., Liu, W., Yang, L., Sun, S., Hu, W., Zhang, F., & Li, W. (2018). Deepunet: a deep fully convolutional network for pixel-level sea-land segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(11), 3954–3962.

Lin, H., Shi, Z., & Zou, Z. (2017). Maritime semantic labeling of optical remote sensing images with multi-scale fully convolutional network. *Remote sensing*, *9*(5), 480.

Lin, H., Zhao, C.-s., & Shu, N. (2003). Edge detection based on canny operator and evaluation. *Journal of Heilongjiang Institute of Technology*, *2*, 3–6.

Liu, H., & Jezek, K. (2004). Automated extraction of coastline from satellite imagery by integrating canny edge detection and locally adaptive thresholding methods. *International journal of remote sensing*, *25*(5), 937–958.

Liu, L., Chen, G., Pan, Z., Lei, B., & An, Q. (2018). Inshore ship detection in sar images based on deep neural networks. In *Igarss 2018-2018 ieee international geoscience and remote sensing symposium* (pp. 25–28).

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 3431–3440).

Ouma, Y. O., & Tateishi, R. (2006). A water index for rapid mapping of shoreline changes of five east african rift valley lakes: an empirical analysis using landsat tm and etm+ data. *International Journal of Remote Sensing*, *27*(15), 3153–3181.

Pashaei, M., & Starek, M. J. (2019). Fully convolutional neural network for land cover mapping in a coastal wetland with hyperspatial uas imagery. In *Igarss 2019-2019 ieee international geoscience and remote sensing symposium* (pp. 6106–6109).

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 234–241).

Shusterman, E., & Feder, M. (1994). Image compression via improved quadtree decomposition algorithms. *IEEE Transactions on Image Processing*, *3*(2), 207–215.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sui, B., Jiang, T., Zhang, Z., Pan, X., & Liu, C. (2020). A modeling method for automatic extraction of offshore aquaculture zones based on semantic segmentation. *ISPRS International Journal of Geo-Information*, *9*(3), 145.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first aaai conference on artificial intelligence.*

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1–9).

Tms rule @ONLINE. (n.d.). Retrieved from `https://wiki.osgeo.org/wiki/Tile\_Map\_Service\_Specification`

Wang, N., Chen, F., Yu, B., & Qin, Y. (2020). Segmentation of large-scale remotely sensed images on a spark platform: A strategy for handling massive image tiles with the mapreduce model. *ISPRS Journal of Photogrammetry and Remote Sensing*, *162*, 137–147.

Wei, Y., & He, Y. (2016). Automatic water line detection for an usv system. In *2016 ieee international conference on cyber technology in automation, control, and intelligent systems (cyber)* (pp. 261–266).

Xiaohua, Z., Yunlong, C., & Xiuchun, Y. (2004). On fractal dimensions of china's coastlines. *Mathematical Geology*, *36*(4), 447–461.

Yang, X., & Liu, Z. (2005). Quantifying landscape pattern and its change in an estuarine watershed using satellite imagery and landscape metrics. *International Journal of Remote Sensing*, *26*(23), 5297–5323.

Yu, B., Yang, L., & Chen, F. (2018). Semantic segmentation for high spatial resolution remote sensing images based on convolution neural network and pyramid pooling module. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(9), 3252–3261.

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2881–2890).

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 3–11). Springer.

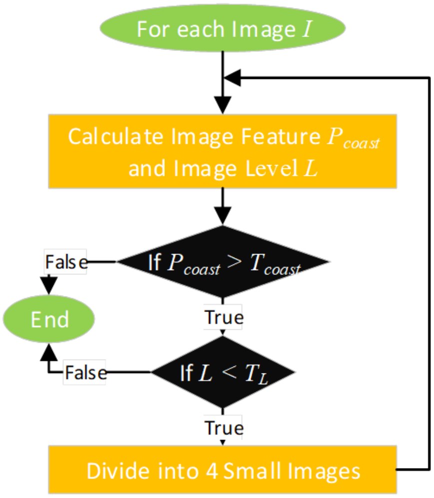

Figure S1. Process of quadtree decomposition

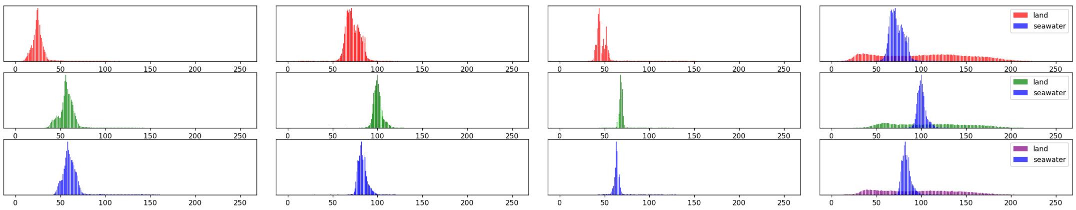

Figure S2. Histogram of waterbody of each band at (a)level12 (b)level14 (c)level16, and (d) comparison of histogram between waterbody and land.

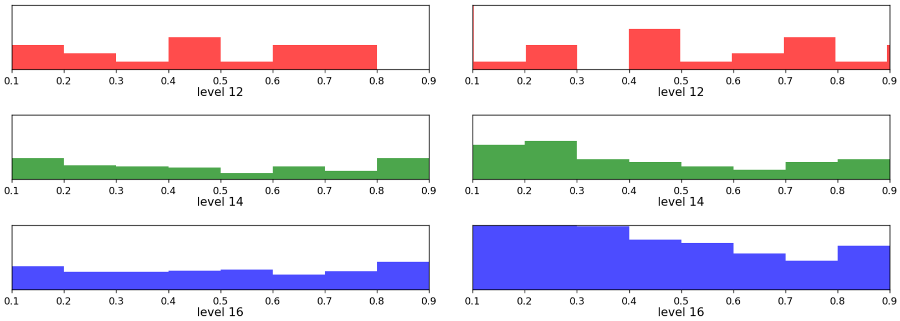

Figure S3. Statistics of proportion of seawater area and freshwater in tile images of different scales

Figure S4. Coastline extraction result (a) directly using Unet++, (b) our method and (b) split line of quadtree decomposition.

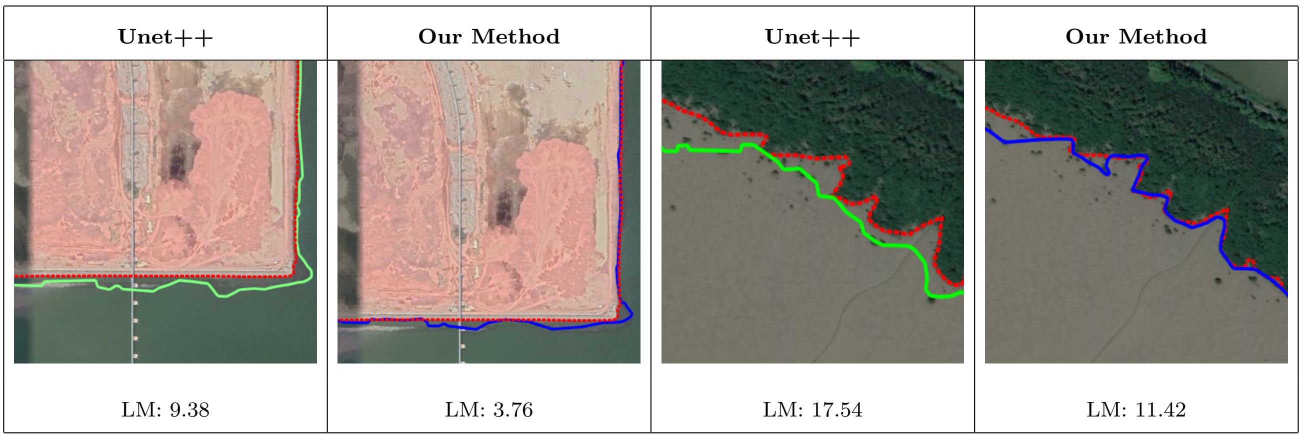

Figure S5. Extraction results of the local area and line matching index

Table S1. Time consumption of each method

Time Consumption	**Identification**	**Segmentation**	**Total**
Unet++		920.82s	920.82s
Our Method	44.13s	60.24s	104.37s
Number of Tiles	**Identification**	**Segmentation**	
Unet++		53728	
Our Method	4794	3536	