

Article

Sea–Land Segmentation Using HED-UNET for Monitoring Kaohsiung Port

Shih-Huan Tseng *  and Wei-Hao Sun

Department of Computer and Communication Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 82245, Taiwan

* Correspondence: shtseng@nkust.edu.tw

Abstract: In recent years, it has become a trend to analyze shoreline changes through satellite images in coastal engineering research. The results of sea–land segmentation are very important for shoreline detection. CoastSat is a time-series shoreline detection system that uses an artificial neural network (ANN) on sea–land segmentation. However, the method of CoastSat only uses the spectral features of a single pixel and ignores the local relationships of adjacent pixels. This impedes optimal category prediction, particularly considering interference by climate features such as clouds, shadows, and waves. It is easy to cause the classifier to be disturbed in the classification results, resulting in classification errors. To solve the problem of misclassification of sea–land segmentation caused by climate interference, this paper applies HED-UNet to the image dataset obtained from CoastSat and learns the relationship between adjacent pixels by training the deep network architecture, thereby improving the results of erroneous sea–land segmentation due to climate disturbances. By using different optimizers and loss functions in the HED-UNet model, the experiment verifies that Adam + Focal loss has the best performance. The results also show that the deep learning model, HED-UNet, can effectively improve the accuracy of the sea–land segmentation to 97% in a situation with interference from atmospheric factors such as clouds and waves.

Keywords: sea–land segmentation; deep learning; satellite images

MSC: 68T07



Citation: Tseng, S.-H.; Sun, W.-H. Sea–Land Segmentation Using HED-UNET for Monitoring Kaohsiung Port. *Mathematics* **2022**, *10*, 4202. <https://doi.org/10.3390/math10224202>

Academic Editor: Bo-Hao Chen

Received: 28 September 2022

Accepted: 2 November 2022

Published: 10 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

It is crucial to observe and quantify changes along the shoreline to grasp the current dynamic changes and long-term trends of coastal changes and to provide key reference information for relevant government agencies in long-term coastal planning and management and coastal disaster prevention.

In recent years, satellite imagery has become easier to obtain than ever, which has saved time and manpower. Therefore, the study of shoreline changes focuses on satellite image analysis. Our previous work [1] explored the sea–land segmentation methods in the CoastSat [2] detection system proposed by Vos et al. and compared the classification results and accuracy of various classifiers for the sea–land segmentation, including artificial neural network (ANN), k-nearest neighbors (KNN), decision tree (DTC), linear support vector machine, and nonlinear support vector machine. However, the method of CoastSat only uses the spectral features of a single pixel and ignores the local relationships of adjacent pixels, making it difficult to give the best category prediction, and the interferences to the climate include: clouds, fog, waves, etc. It is easy to cause the classifier to be disturbed in the classification results, resulting in classification errors. Our motivation is to solve classification errors as shown in Figure 1. The results obtained by the classifier for sea–land segmentation contain errors attributable to (a) waves, (b) clouds, and (c) fog. Therefore, the contributions of this paper are as follows:

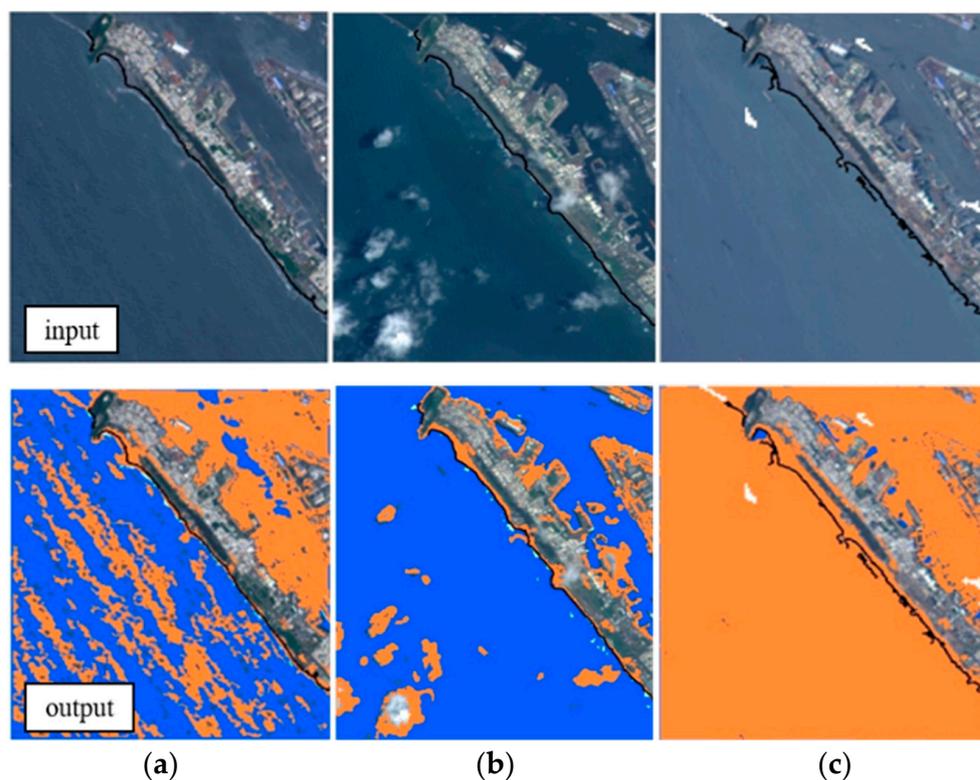


Figure 1. Misclassification results of sea–land segmentation on the Kaohsiung Port of (a) waves, (b) clouds, and (c) fog.

1. Our work applies the HED-Unet [3] to improve the segmentation of sea and land in the factor interferences of the atmosphere such as clouds and waves.
2. We have collected our own satellite imagery of important ports in Taiwan with climate influences.

The rest of this paper is organized as follows: Section 2 introduces the related work for sea–land segmentation using deep learning. Section 3 describes the steps in sea–land segmentation from satellite images using HED-Unet, including data collecting, data labeling, and model training. Section 4 presents the experiment results intended to evaluate the performance of the proposed methods and the limitations of this work are discussed in Section 4. Section 5 concludes the paper and outlines future work.

2. Related Work

In the research topic of sea–land segmentation using satellite images, many studies have replaced conventional learning methods with deep learning due to the rapid development and wide application of deep learning in recent years [4]. Dongcai et al. proposed SeNet [5] and mentioned that there are two problems in using DeconvNet [6] directly for sea–land segmentation. First, more complex texture and intensity variations appear in inland areas. Both conventional methods and DeconvNet produce misclassifications in land areas, and secondly, they perform [5] poorly on some slender structures. For the above reasons, this study makes two innovations in the network structures. First, a local positive normalization is proposed to obtain better spatially consistent results, getting rid of complex morphological operations commonly used in traditional methods, and second, segmentation and edge detection results are simultaneously obtained using a multi-task loss. An additional structured edge detection branch can further refine the segmentation results and significantly improve edge accuracy, and the system architecture diagram is shown in Figure 2.

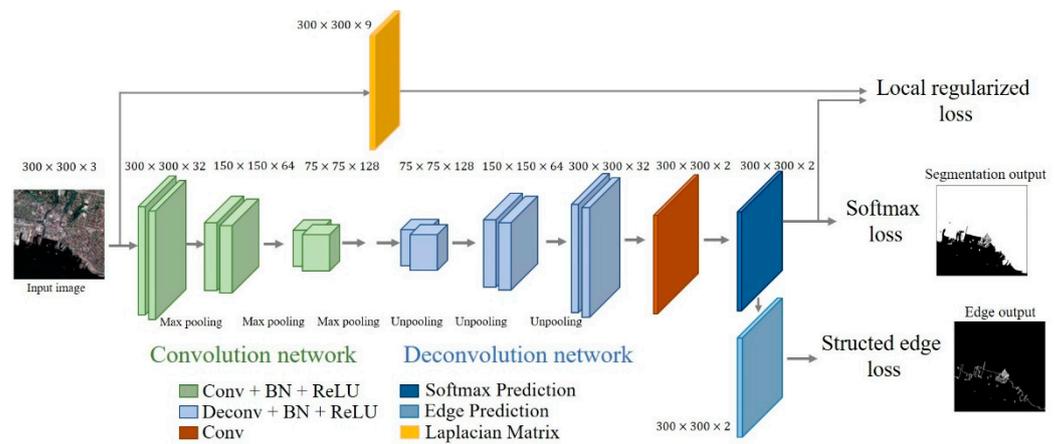


Figure 2. SeNet system architecture [5].

However, SeNet is easily affected by disturbances such as waves and shadows. The reason is that its architecture cannot provide a deeper convolutional layer for the local relationship between adjacent pixels, so it cannot clearly distinguish detailed features in the classification of coastal and land boundaries, such as ships, ports, vegetation, etc. Another convolutional neural network structure commonly used for segmentation tasks is U-Net [7], which is used for biomedical image segmentation. Its architecture consists of a shrinking path and an expanding path, and its feature maps are cropped and copied from the shrinking path for corresponding upsampling in the expanding path.

DeepUnet [8], proposed by Ruirui Li et al., performed the best at the time. Satellite and medical images differ in the occurrence of small objects in the images. Deeper convolutional neural networks are required to consider global contention and local features. In this research, inspired by the U-Net architecture, two connections are proposed in the network model, namely, U-connection and connection, to reduce the loss of wrong information and accelerate the collection speed. After increasing the number of layers in the experimental results, it is better than Unet and SeNet, it has successfully improved the problems such as wave interference, and it has a more accurate effect on the boundary classification of sea and land. The system architecture diagram of this method is shown in Figure 3.

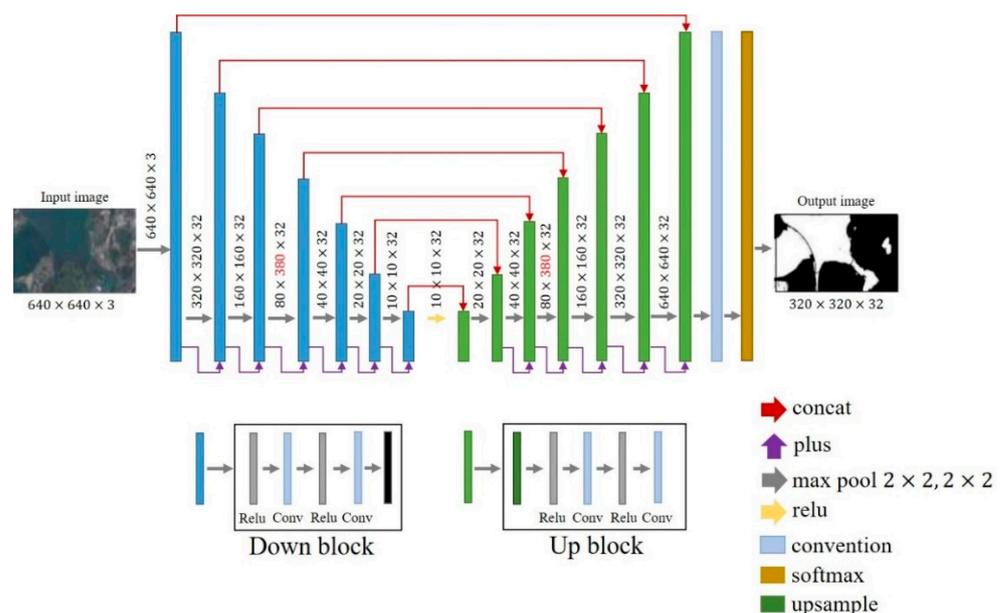


Figure 3. DeepUnet system architecture [8].

Finally, HED-Unet [3] was recently proposed by Konrad Heidler et al. to detect the Antarctic coastline. Encoder–decoder-based works are mainly used in image enhancement to obtain high-resolution images [9–11]. In HED-Unet, the main purpose is to obtain images of different resolutions and to classify images of different resolutions in different regions through the attention mechanism. For example, coastal areas use higher-resolution imagery for detailed classification, while land, ports, buildings, and other areas far from the coast use lower-resolution imagery.

Since the southern and coastal land is white iceberg, it is more difficult to distinguish the sea surface from the land. In the experiment, many existing deep learning models cannot accurately predict the Antarctic coast, but HED-Unet introduces the idea that humans will perform segmentation and edge detection at the same time when observing edges. So, the Unet [7] semantic segmentation is combined with the holistically-nested edge detection (HED) [12] model architecture. Training efficiency can be improved by deep supervision [13] for side predictions at multiple resolutions. Finally, a hierarchical attention mechanism [14] is introduced to adaptively incorporate these multiscale predictions into the final model output. Experiments show that the dataset covers part of the Antarctic coast and outperforms DeepUnet. Figure 4 shows the system architecture diagram.

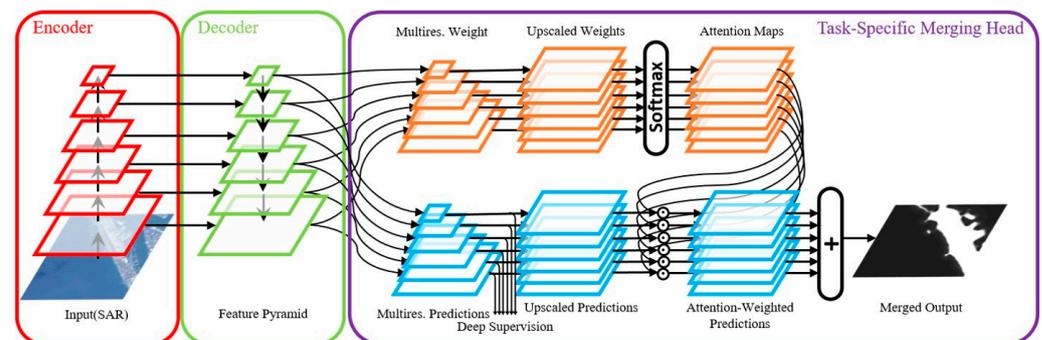


Figure 4. HED-Unet architecture with only the segmentation head [3].

In our previous work [1], there were waves and clouds in the satellite image, and these noises were classified as land by the machine learning model. The CoastSat system very easily confuses the boundaries of beaches and waves in terms of pixel intensity. The difference between pixel intensities is not considered when the CoastSat system trains the model, while the HED model can encode deep feature maps with rich adjacent relationship information into shallower feature maps while gaining the advantage of deep and shallow features.

For the above reasons, the novelty of our work is making the model have a better anti-climate-interference ability. In order to do this, the model must take into consideration the feature relationships between adjacent pixels in satellite images, and HED-Unet extracts features from images of different scales, and thus can obtain the relationship between adjacent pixels. Therefore, we use HED-Unet as a model to improve the results of sea–land segmentation.

3. Sea–Land Segmentation from Satellite Images Using HED-Unet

We used HED-Unet [3] as a deep learning model to improve the impact of climate disturbance. The method of data collection utilizes Google Earth Engine [15] to obtain the same satellite images as in CoastSat. Different from the CoastSat, our work did not use spectral images for feature extraction and used RGB satellite images instead. There are also differences in the methods of data labeling and model training.

In the system process, the satellite image of Kaohsiung Port is obtained first through Google Earth Engine (GEE) API, and then the satellite image is labeled through LabelMe [16], and the annotation is divided into two categories, seawater and land. The satellite image is used as the input to the deep learning model HED-Unet to perform feature

extraction and prediction of images of different scales and then fuse the predictions of different scales to finally obtain sea–land segmentation images. The high-level structure of our work is shown in Figure 5.

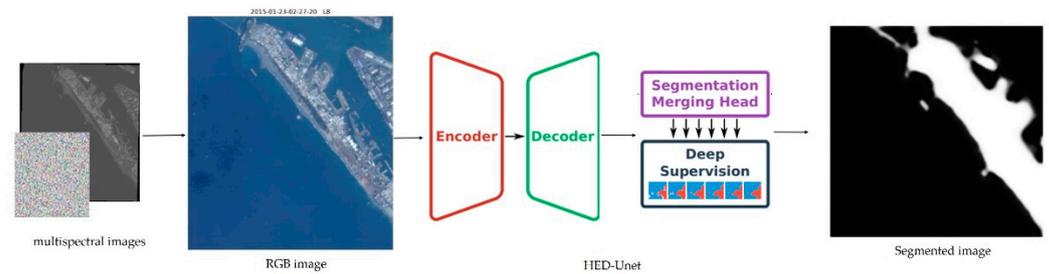


Figure 5. High-level structure of sea–land segmentation system with HED-Unet [3].

3.1. Data Collecting

The first step of the method is to use Google Earth Pro to frame the target to obtain the RGB image of the coast from the satellite’s multi-spectral image. The frame method is shown in Figure 6. A total of five points are marked, and the start and end points will be clockwise. We then overlap, frame the target, and convert it through Google Earth Pro to obtain a .kml file, which records the geographic information and coordinates of the target. Then, convert the kml file into latitude and longitude coordinates and use the time and satellite mission (Landsat) as input parameters to the Google Earth Engine. We follow the steps of CoastSat [2] to obtain images of relevant spectral bands, as shown in Figure 7, including multispectral band images and panchromatic, which sharpens the judgment image.

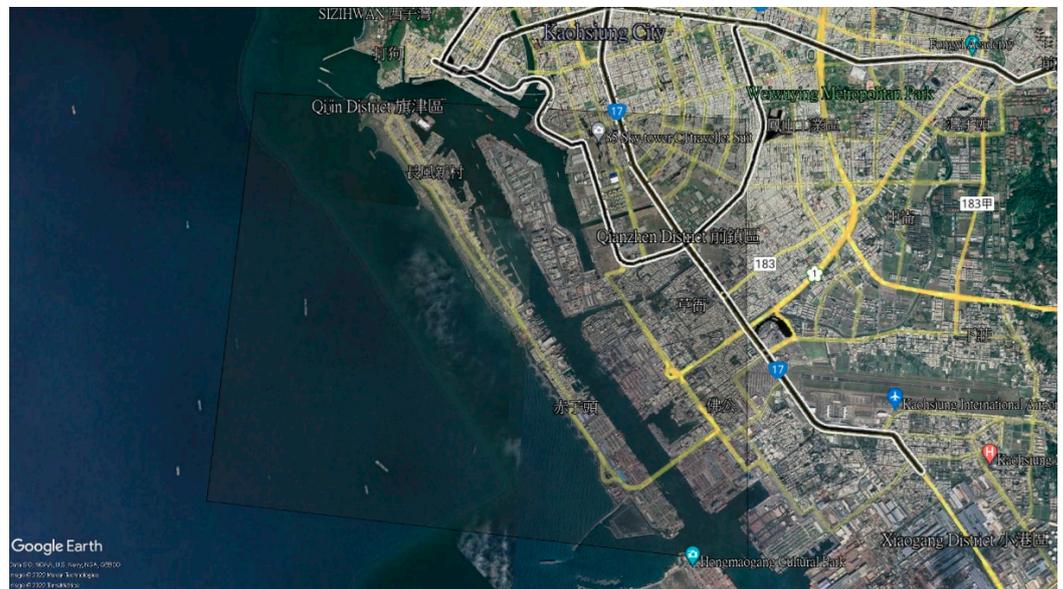


Figure 6. Use Google Earth Pro to select the target coast. Image©2022 Maxar Technologies, Image©2022 TerraMetrics.

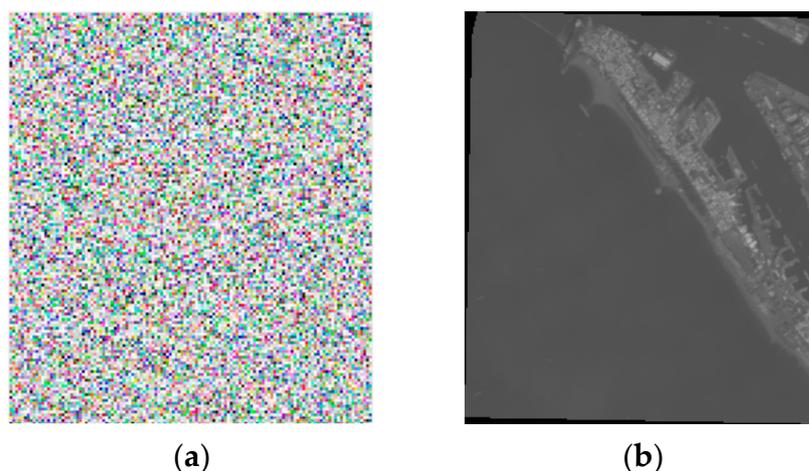


Figure 7. Landsat spectral images: (a) Multispectral band; (b) Panchromatic sharpening band.

The final step is to process the multispectral images to obtain RGB images. To improve the efficiency and accuracy of sea–land segmentation, the methods use panchromatic image sharpening and downsampling to enhance the resolution and obtain better coastline-detection performance; and cloud masking to reduce the misjudgment of white foam during subsequent classification. Figure 8a is the pre-processed color satellite image of the Kaohsiung port, and Figure 8b is an image of the Kaohsiung port covered by clouds and fog.

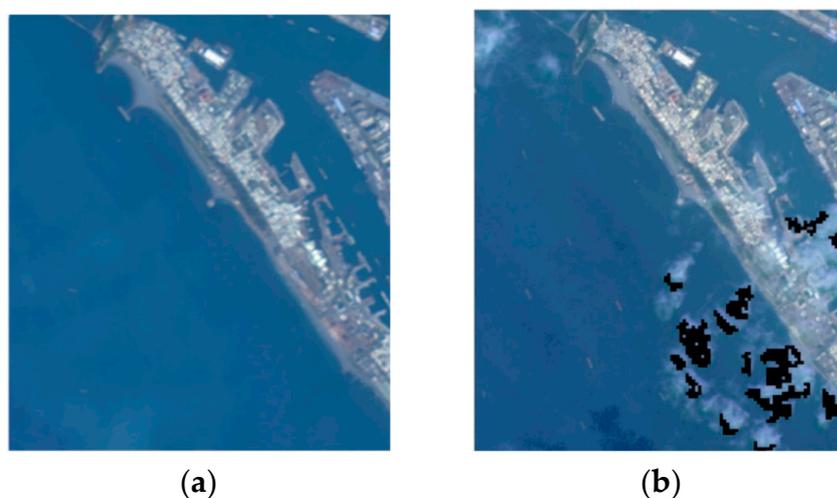


Figure 8. Kaohsiung Port: (a) Color image; (b) Cloud removal image.

3.2. Data Labeling

RGB images are obtained after processing the multispectral images. Next, the RGB images need to be labeled. The labeling method is carried out through the LabelMe tool. LabelMe [16] was made by the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL) for image annotation. The Annotation Suite, which allows users to create custom annotation tasks or perform image annotation, has the advantage of being easy to use.

The satellite image is framed to mark the target through the point, and the target is land. During the labeling process, it was found that the unlabeled part is regarded as the background and also as a category. Therefore, in this study, the land is the foreground, and the sea is the background. In the process, the Google Maps satellite imagery is used for labeling and comparison to ensure that there is no wrong labeling during the labeling process. The labeling process is shown in Figure 9a. During the process, images of different scales of the Kaohsiung seaport are also labeled to ensure thorough training.

The robustness of each scale includes near, medium, and far, and the labeling results are shown in Figure 9b–d. The testing data are the satellite images of Taichung and Keelung ports, as shown in Figure 10. The reason for choosing Kaohsiung Port is that this paper was supported by developing a school-specific project of the National Kaohsiung University of Science and Technology. The Kaohsiung port, Taichung port, and Keelung port are important ports in Taiwan. All of them have encountered cloud, fog, and wave interferences in sea–land segmentation.

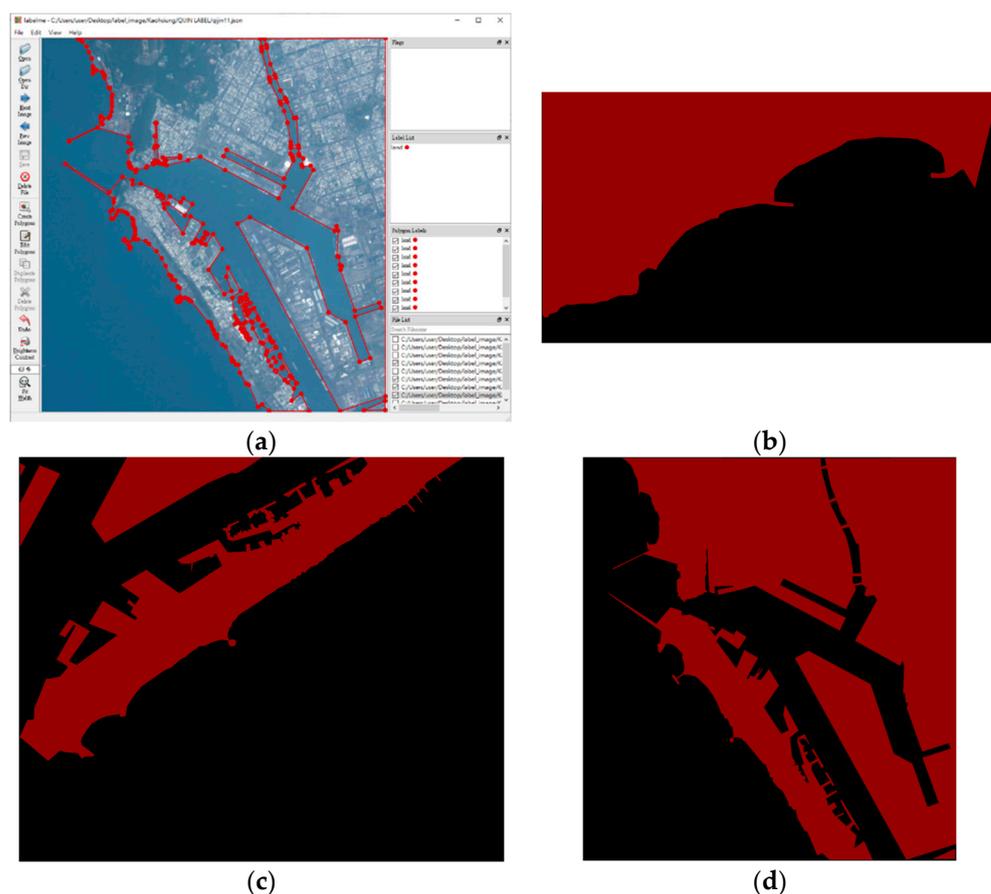


Figure 9. Labeling process and results of Kaohsiung Port: (a) LabelMe interface; (b) Near scale; (c) Medium scale; (d) Far scale.



Figure 10. Labeling results of testing data: (a) Taichung Port; (b) Keelung Port.

3.3. Model Training

After the process of data collection and data annotation, we input the acquired images into the model for training. We divide it into three parts for detailed description, which are the details and process of image input pre-processing, feature extraction, and feature fusion. First, the input image is cut into 256*256 tiles as shown in Figure 11, and each tile is converted into a vector as the input of the model structure. This method can be carried out in the case of scarce labeled data.

Second, feature extraction is performed through Unet to obtain context information and location information. Since the process is up-sampling and down-sampling, it can also be called the process of encoding and decoding. The advantage lies in that feature fusion is performed through channels. Avoid missing feature information during up-sampling. In the second section of the second chapter, it is mentioned that five steps of up-sampling and down-sampling are used in the system, and six information features of different scales are obtained to form a feature pyramid; the features of different scales represent the meaning of shallow features and are used to represent textures, and deep low-resolution features are used to determine the relationship between adjacent pixels.

Finally, in the process of training the HED model, Deep Supervision [13] is used to encode the deep feature map with rich adjacent relationship information to the shallower feature map to obtain the advantages of deep and shallow features at the same time. The process is to first predict the feature map in Deep Supervision to obtain the probability distribution map of each layer's feature map and its classification, and then compare the predicted result with the target to obtain additional loss values as the basis for additional judgment, as shown in Figure 11.

In the network architecture, merging heads combine information of different scales through convolution. The purpose is to obtain the required features of each region through fusion according to the different features required by different regions in the satellite image. In the study [3], the authors propose to use an attention mechanism for fusion, which allows the network to focus on the features it thinks are most useful for each pixel of the current scene, rather than setting fixed weights for feature fusion. Therefore, the attention mechanism allows the model to pay attention to resolutions at different scales.

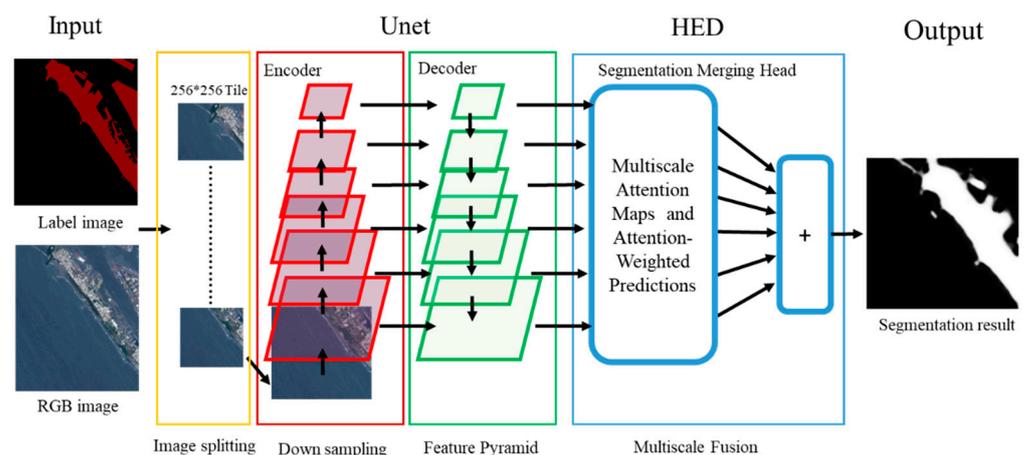


Figure 11. Flow chart of sea-land segmentation training with HED-Unet.

4. Experiments and Results

4.1. Experiment Setting

The experimental environment of this research is mainly based on personal computers, and the training is carried out on the local machine, including the software and hardware equipment used in the experimental process, as shown in Table 1.

Table 1. Experimental environment.

Component	Specifications
Central Processing Unit (CPU)	Intel Core I7-8700
Graphics Processing Unit (GPU)	NVIDIA GeForce GTX 950M
Compute Unified Device Architecture (CUDA)	CUDA 10.2
Random Access Memory (RAM)	64GB
Operating System	Windows 10
Programming Language	Python 3.7
Packages	Google Earth Engine-API 1.12.8 LabelMe 4.5.9 Pytorch 1.9.0

HED-Unet deep learning is used to set the parameters used in the experimental design for model training. During the training process of this experimental design, eight satellite images of Kaohsiung Port are used for data annotation. The images include different situations such as clouds, fog, and waves; they are cut into 3000 tiles for training through image processing; and they use Taichung Port and Keelung Port as test locations in the test data set. The test sets are Keelung Port and Taichung Port. There are eight images individually and each image is divided into 3000 tiles individually. In the experimental design, we use different optimizers including Adam and stochastic gradient descent (SGD). For comparison, we also use different loss functions including binary cross-entropy (BCE) and focal loss, and in the learning depth of Unet, as mentioned in the literature, more than five layers cannot increase the performance of prediction, so five layers were selected, in terms of parameters. As shown in Table 2, the parameters are set in the design of this training.

Table 2. Parameter settings of HED-Unet.

Input channels	3
Output channels	2
Stack height	5 (literature mentioned)
Batch size	20 (limited by computer equipment)
Epochs	50 (limited by computer equipment)
Learning rate	0.001 (Make sure to converge)
Loss function	Binary Cross-Entropy (BCE) [17], Focal Loss [18]
Active function	Sigmoid
Optimizer	Adam [19], Stochastic Gradient Decent (SGD) [20]

4.2. Results

As the optimizer in the model, Adam is an adaptive algorithm that is suitable for unstable functions and gradient noise problems and can have different learning rates for different memory calculations. Zhang et al. (2019) proved that Adam outperforms the SGD for the attention model [21]. Stochastic gradient descent (SGD) [20] is a typical optimizer algorithm in the training model. It is fast for large data sets, but the learning rate cannot be adjusted in the process, and it is easy to converge to a locally optimal solution. The results also show differences in their predictions. Focal loss [18] is a loss function proposed by Tsung-Yi Lin et al. in 2017. Focal loss is proposed for object detection to solve extreme foreground–background class imbalance encountered during the training of dense detectors as the central cause. So, there will be an extreme imbalance in the

calculation of loss, and this loss function is mainly used to solve the category imbalance in the classification problem. Its effect can be seen in the experimental results.

Figure 12 shows the results of the sea–land segmentation experiment performed by HED-UNet with different loss functions and optimizers; Figure 12a is the input satellite image; Figure 12d is the real value of sea–land segmentation obtained in the data labeling; and Figure 12b is the use of the Adam optimizer with the binary cross-entropy (BCE) loss function. For the predicted results of the sea–land segmentation, Figure 12e is the predicted result using the Adam optimizer and the loss function of focal loss; Figure 12c is the predicted result using the SGD optimizer with the binary cross-entropy loss function; and Figure 12f is the result predicted using the SGD optimizer with the loss function of focal loss. Although the segmentation of sea and land is more accurate in Figure 12f, Figure 12c is clearer in terms of clarity, and it used the SGD optimizer as well. Although it can be seen that some predictions work well locally, it is less clear than using the Adam optimizer as a whole.

Accuracy is examined in the evaluation criteria, and we use the verification data and test data for different locations to conduct objective verification. In the experimental results in Table 3, it is found that the accuracy of using focal loss as the loss function is quite high. The convergence loss value is also close to 0, and the SGD optimizer with the BCE loss function has the worst performance. The accuracy rate of the test data is only 0.597. There are differences between the HED-UNet and the CoastSat system in the training model. As HED-UNet cuts satellite images and labeled images into multiple partial images through image preprocessing data, it is less likely to be disturbed by other features during the training process, and its categories are only ocean and land. Meanwhile, there are four categories in each classification model in the CoastSat system, which is less prone to misclassification.

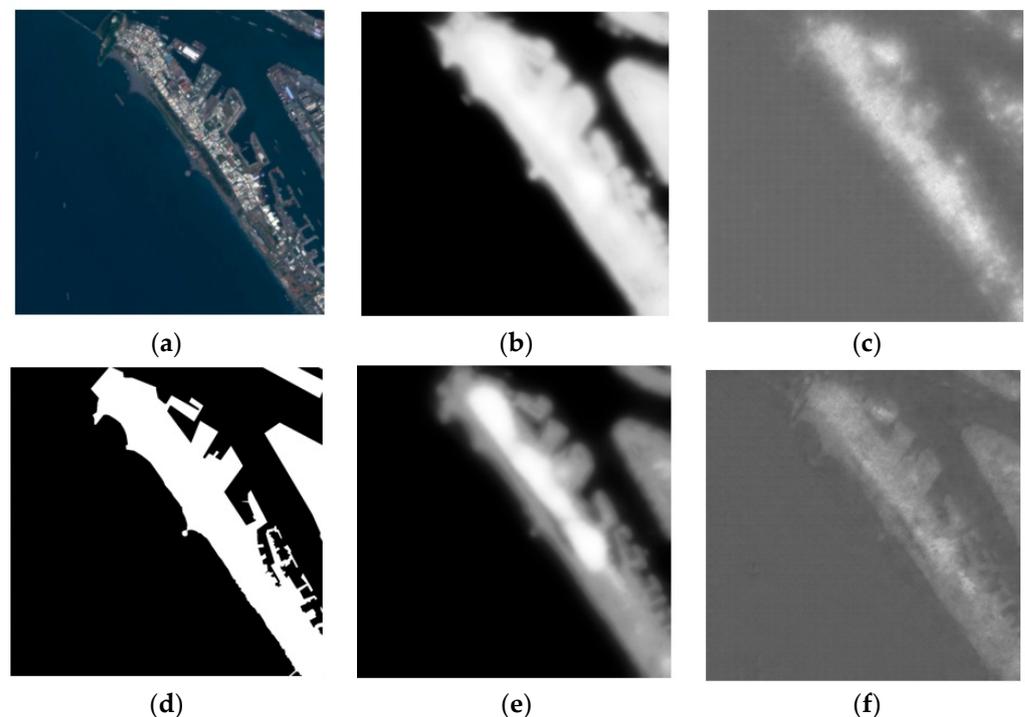


Figure 12. Sea–land segmentation results with different loss functions and optimizers: (a) RGB image; (b) BCE + Adam; (c) BCE + SGD; (d) Ground truth; (e) Focal loss + Adam; and (f) Focal loss + SGD.

Table 3. Accuracy of sea-land segmentation using HED-Unet with different loss functions and optimizers.

Model	Adam+ BCE	Adam+ Focal	SGD+ BCE	SGD+ Focal
Val Acc.	0.933	0.972	0.720	0.956
Test Acc.	0.915	0.983	0.597	0.931

In addition to the analysis of the accuracy, the change of the loss value within 50 epochs is also observed according to the use of different optimizers and different loss functions. The lesser the loss value, the greater the convergence effect. As shown in 13a, based on the change of the loss value of the Adam optimizer with the binary cross-entropy loss function, the convergence starts at the third epoch during the training process, and the final convergence loss value is about 0.17. As shown in Figure 13b, based on the change of the loss value using the Adam optimizer and the loss function of focal loss, the convergence starts at the sixth epoch during the training process, and the final convergence loss value is about 0.15. As shown in Figure 13c, based on the change of the loss value using the SGD optimizer with the binary cross-entropy loss function, the convergence starts at the fifth epoch during the training process, and the final convergence loss value is about 0.48. As shown in Figure 13d, based on the change of the loss value using the SGD optimizer with the focal loss, the convergence starts at the second epoch during the training process, and the final convergence loss value is about 0.25.

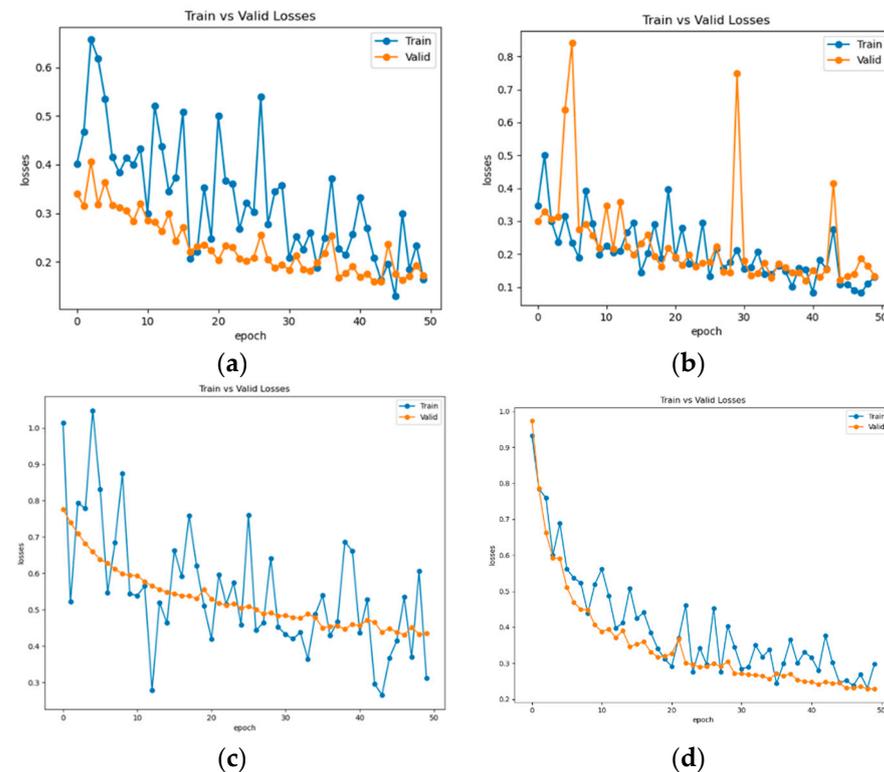


Figure 13. Training and validation loss values with different loss functions and optimizers: (a) Adam + BCE; (b) Adam + Focal loss; (c) SGD + BCE; and (d) SGD + Focal loss.

Based on the use of the Adam optimizer and focal loss, the convergence value is at least 0.15, and the convergence value of the SGD optimizer with binary cross-entropy loss function is about 0.5. The focal loss is more stable in convergence for different optimizations. It has little effect, and the convergence value is about 0.2. According to its convergence

results, it can be found that using the Adam optimizer with the focal loss can achieve a smaller loss value on the training model.

Next, it is explained that the ANN in Coastsat and HED-Unet in our work have differences in data labeling in the experimental method. Coastsat uses pixel sampling for local labeling, while HED-Unet performs global labeling for the entire satellite image while training the model line. The method of feature extraction is also different. In the ANN classification model in Coastsat, the feature extraction method obtains 20 features in each sampled pixel through multispectral image information and uses these features to develop four categories. In the training of the HED-Unet model, feature extraction uses encoding and decoding to obtain feature maps of different scales, which can be used to predict satellite images of different scales.

Finally, we will explain and compare the results of the sea–land segmentation caused by different climate disturbances in sequence, as shown in Figure 14a–c. Because of the different conditions caused by climate influence, the Kaohsiung port includes waves, fog, and clouds, it can be seen from the figures that the pixel-based classification lacks information on the surrounding relationships with the adjacent pixels, resulting in the wrong classification of the land and the sea surface. In addition, because the Kaohsiung port has a black sand beach, it is easy to detect areas with similar pixel intensities. However, the HED-Unet model can better grasp the adjacent relationships due to the increase in the number of layers. It can be seen from the results that the sea and land segmentation results of the Kaohsiung port in different situations are not greatly affected, but the results in the details of the port classification need to be strengthened.

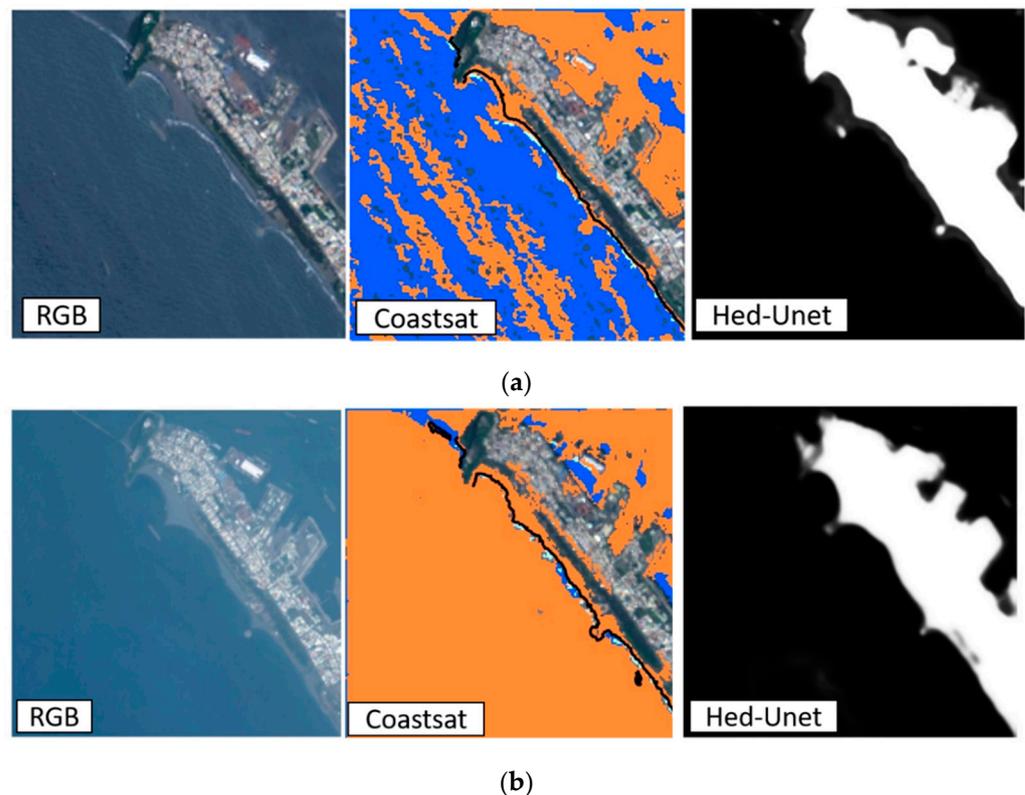


Figure 14. Cont.

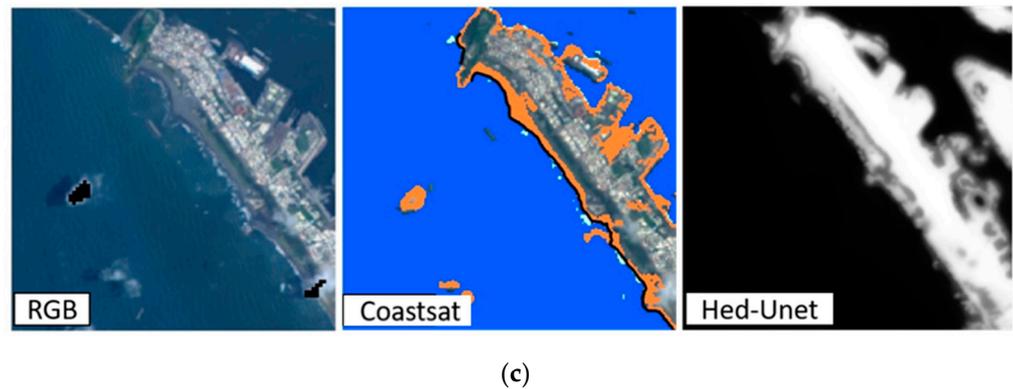


Figure 14. Comparison of sea–land segmentation results under atmospheric interference (a) waves (b) fog (c) clouds.

4.3. Discussion

This study conducts experiments through data collection, data labeling, and training models. In the process of using the deep learning model HED-Unet on the satellite images of the Kaohsiung port, one of the limitations in our work lies in data labeling. In the data labeling process, satellite images are occluded by a large number of clouds and fog. The data annotation of them cannot be effectively performed in the sea–land segmentation task.

Before training the model, it is necessary to manually screen images with large numbers of clouds and fog, which takes a lot of manpower and time.

The other limitation, as illustrated by the results in Figure 15, is that the task of sea–land segmentation cannot be effectively performed on large-scale images occluded by clouds and fog. The reason is that Figure 15's land features and sea surface features are heavily occluded by clouds and fog, which makes the model unable to obtain effective features for training during the training process. It is hoped that in the future, continuous time-series satellite image data sets can be obtained, and the time series model architecture can be used for prediction. It is expected that a large number of cloud and fog masks can also detect its sea and land segmentation results.

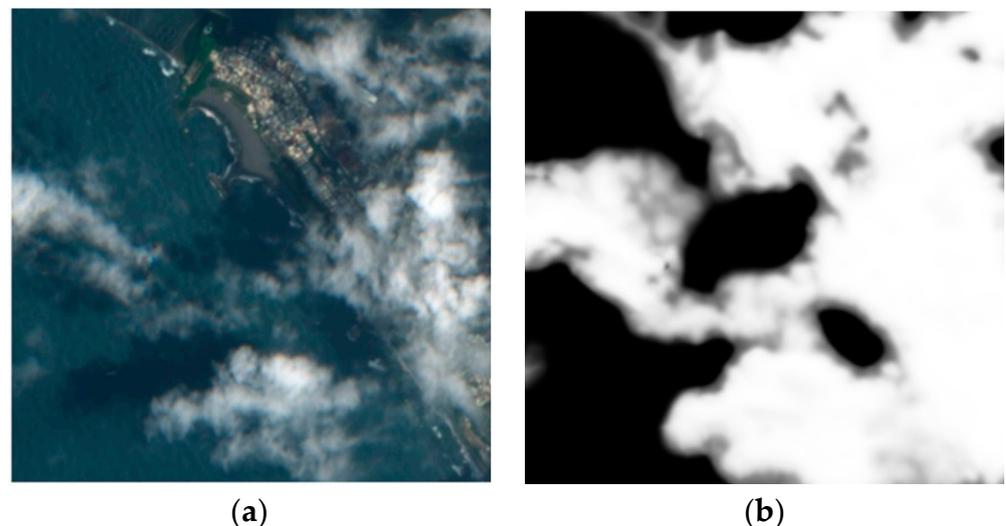


Figure 15. Sea–land segmentation results are interfered with by large areas of cloud and fog. (a) Satellite image (b) Classification result.

5. Conclusions

In recent years, rapid climate change and global warming have made it very important to study the changes in the coastlines of ports. In this study, the Coastsat shoreline detection

system was used to obtain satellite images, and then the images were segmented into sea and land.

This study uses the deep learning model HED-Unet to perform sea–land segmentation on the satellite images of Kaohsiung Port to improve the classifier against the interference of atmospheric factors such as clouds and waves. The experimental results of HED-Unet in the sea–land segmentation task show that the use of the Adam optimizer and the loss function of focal loss has the best convergence and prediction results. It is suitable for the general sea surface, and the limit of the deep learning model HED-Unet was also found during the experiment; in the case of a large number of occlusions by clouds and fog, its prediction results are not ideal.

Finally, we hope that the contribution of this research can give some feedback to coastal engineering. We also hope that in future experiments of various classification models, more satellite images of ports around Taiwan can be added to the data set to increase the objectivity of the experiment. The problem would be best addressed by implementing a time series model architecture.

Author Contributions: Conceptualization, S.-H.T.; methodology, S.-H.T. and W.-H.S.; software, W.-H.S.; validation, S.-H.T. and W.-H.S.; formal analysis, S.-H.T.; investigation, W.-H.S.; resources, W.-H.S.; data curation, W.-H.S.; writing—original draft preparation, W.-H.S.; writing—review and editing, S.-H.T.; visualization, W.-H.S.; supervision, S.-H.T. All authors have read and agreed to the published version of the manuscript.

Funding: The Ministry of Education, Taiwan, and project name is Higher Education SPROUT Project.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wei-Hao, S.; Tseng, S.H. Comparisons of Classification Models on COASTSAT. In Proceedings of the 2021 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Hualien City, Taiwan, 16–19 November 2021; pp. 1–2.
2. Vos, K.; Splinter, K.D.; Harley, M.D.; Simmons, J.A.; Turner, I.L. CoastSat: A Google Earth Engine-enabled Python toolkit to extract shorelines from publicly available satellite imagery. *Environ. Model. Softw.* **2019**, *122*, 104528. [\[CrossRef\]](#)
3. Heidler, K.; Mou, L.; Baumhoer, C.; Dietz, A.; Zhu, X.X. HED-UNet: Combined segmentation and edge detection for monitoring the Antarctic coastline. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 4300514. [\[CrossRef\]](#)
4. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [\[CrossRef\]](#)
5. Cheng, D.; Meng, G.; Cheng, G.; Pan, C. SeNet: Structured edge network for sea–land segmentation. *IEEE Geosci. Remote Sens. Lett.* **2016**, *14*, 247–251. [\[CrossRef\]](#)
6. Noh, H.; Hong, S.; Han, B. Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1520–1528.
7. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
8. Li, R.; Liu, W.; Yang, L.; Sun, S.; Hu, W.; Zhang, F.; Li, W. DeepUNet: A deep fully convolutional network for pixel-level sea-land segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3954–3962. [\[CrossRef\]](#)
9. Fang, Y.; Xu, L.; Chen, Y.; Zhou, W.; Wong, A.; Clausi, D.A. A Bayesian Deep Image Prior Downscaling Approach for High-Resolution Soil Moisture Estimation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4571–4582. [\[CrossRef\]](#)
10. Liu, K.; Liu, D.; Li, L.; Yan, N.; Li, H. Semantics-to-signal scalable image compression with learned revertible representations. *Int. J. Comput. Vis.* **2021**, *129*, 2605–2621. [\[CrossRef\]](#)
11. Xu, J.; Zhou, W.; Chen, Z.; Ling, S.; Le Callet, P. Binocular rivalry oriented predictive autoencoding network for blind stereoscopic image quality measurement. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 5001413. [\[CrossRef\]](#)
12. Xie, S.; Tu, Z. Holistically-nested edge detection. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1395–1403.
13. Lee, C.-Y.; Xie, S.; Gallagher, P.; Zhang, Z.; Tu, Z. Deeply-supervised nets. In Proceedings of the Artificial Intelligence and Statistics, San Diego, CA, USA, 9–12 May 2015; pp. 562–570.

14. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010.
15. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [[CrossRef](#)]
16. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A database and web-based tool for image annotation. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [[CrossRef](#)]
17. Hwang, J.-J.; Liu, T.-L. Pixel-wise deep learning for contour detection. *arXiv* **2015**, arXiv:1504.01989.
18. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
19. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
20. Robbins, H.; Monro, S. A stochastic approximation method. *Ann. Math. Stat.* **1951**, *22*, 400–407. [[CrossRef](#)]
21. Zhang, J.; Karimireddy, S.P.; Veit, A.; Kim, S.; Reddi, S.J.; Kumar, S.; Sra, S. Why are adaptive methods good for attention models? *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 15383–15393.