Article

Combining Segmentation Network and Nonsubsampled Contourlet Transform for Automatic Marine Raft Aquaculture Area Extraction from Sentinel-1 Images

Yi Zhang ^{1,2}, Chengyi Wang ^{1,*}, Yuan Ji ³, Jingbo Chen ¹, Yupeng Deng ^{1,2}, Jing Chen ^{1,2} and Yongshi Jie ^{1,2}

¹ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China; zhangyi@radi.ac.cn (Y.Z.); chenjb@aircas.ac.cn (J.C.); dengyp@radi.ac.cn (Y.D.); chenjing185@mails.ucas.ac.cn (J.C.); jieys@radi.ac.cn (Y.J.)

² School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

³ People's Liberation Army 91039 Troop, Beijing 102401, China; jiyuan13@163.com

* Correspondence: wangcy@radi.ac.cn; Tel.: +86-136-9327-5492

Received: 9 November 2020; Accepted: 17 December 2020; Published: 21 December 2020

Abstract: Marine raft aquaculture (MFA) plays an important role in the marine economy and ecosystem. With the characteristics of covering a large area and being sparsely distributed in sea area, MFA monitoring suffers from the low efficiency of field survey and poor data of optical satellite imagery. Synthetic aperture radar (SAR) satellite imagery is currently considered to be an effective data source, while the state-of-the-art methods require manual parameter tuning under the guidance of professional experience. To preclude the limitation, this paper proposes a segmentation network combined with nonsubsampled contourlet transform (NSCT) to extract MFA areas using Sentinel-1 images. The proposed method is highlighted by several improvements based on the feature analysis of MFA. First, the NSCT was applied to enhance the contour and orientation features. Second, multiscale and asymmetric convolutions were introduced to fit the multisize and strip-like features more effectively. Third, both channel and spatial attention modules were adopted in the network architecture to overcome the problems of boundary fuzziness and area incompleteness. Experiments showed that the method can effectively extract marine raft culture areas. Although further research is needed to overcome the problem of interference caused by excessive waves, this paper provides a promising approach for periodical monitoring MFA in a large area with high efficiency and acceptable accuracy.

Keywords: marine raft aquaculture; Sentinel-1; nonsubsampled contourlet transform; semantic segmentation; fully convolutional network

1. Introduction

According to the data from the Food and Agriculture Organization of the United Nations, aquaculture production worldwide has surpassed that of capture fisheries, and it has been steadily increasing year by year [1]. As the primary mean for coastal aquaculture, marine raft aquaculture plays an important role in the development of the global marine economy and has a considerable impact on the global marine ecosystem. For example, the 2008 Yellow Sea shade green tide outbreak induced by marine raft aquaculture had a negative impact on the Olympic sailing event and led to the loss of RMB 1.3 billion. This serious environmental catastrophe stemmed largely from the environmental pressure

caused by the increasing number of marine raft aquaculture areas [2]. The distribution and number of marine raft aquaculture areas reflect the development status of the fishery as well as the quality of the water environment. The monitoring of marine raft aquaculture areas is of great significance for the protection of marine ecosystems and the sustainable use of marine fishery resources.

Marine raft aquaculture covers a wide area, is far from land, and is sparsely distributed. At present, the relevant government departments mainly rely on field surveys to monitor it. As a result of sea conditions, travel means, and weather, it is difficult to use manual inspection to detect illegal areas and keep track of the distribution and quantity of floating rafts in a timely manner. Remote sensing satellites have the capacity to periodically observe wide areas with few ground restrictions, which enables regular and quick monitoring of marine raft aquaculture areas. A combination of remote sensing monitoring technology and field surveys can provide comprehensive and efficient monitoring of marine aquaculture areas, which is conducive to the orderly development of aquaculture and the protection of natural ecology [3].

Raft culture is a form of aquaculture that uses floats and ropes to form rafts on the sea surface that are fixed to the seabed with cables, which have algae or shellfish suspended on slings [4]. As shown in Figure 1, the floating raft is divided into two parts, i.e., above and below water, and the water surface mainly contains floating balls. The structure of the floating raft makes it difficult for passive remote sensing to capture its reflected signal. Limited by imaging modality, it is difficult to accurately describe the marine raft culture area on optical satellite images [5,6]. Additionally, sea environmental elements such as wind and waves as well as fog render it difficult to extract a research target using optical remote sensing imagery. Synthetic aperture radar (SAR) actively emits electromagnetic waves (fixed frequency beams) and collects reflected and back-scattered signals. It is considered to be the best method for monitoring marine environments because it is not affected by the above elements [7]. Therefore, marine raft aquaculture monitoring through SAR images has practical research significance, especially in coastal cities where mariculture is the main economic activity. Timely and effective monitoring of marine raft aquaculture areas can effectively assist in the planning of marine aquaculture resources.

Figure 1. Raft culture. A floating raft has two parts: underwater (a) and water (b) [5].

In recent years, researchers have focused on SAR-based methods for aquaculture area extraction. Chu et al. extracted raft aquaculture areas using various filtering methods and human–computer interaction [8]. Fan et al. proposed a joint sparse representation classification method to construct meaningful texture features of raft aquaculture on the basis of wavelet decomposition and gray-level co-occurrence matrix (GLCM) statistical methods [9]. Hu et al. improved the statistical region merging algorithm for superpixel segmentation and used a fuzzy compactness and separation clustering algorithm to identify raft aquaculture areas from SAR images [10]. Geng et al. extracted raft aquaculture areas by means of weighted fusion classifiers and sparse encoders [11,12]. These methods are efficient in certain regions with the help of professional experience. However, knowledge-intensive

feature engineering always leads to low robustness. The empirical parameter tuning causes the above methods to not work well with different data and in different regions.

The emergence of convolutional neural networks has provided a way to avoid intensive parameter tuning through deep learning and has led to the focus on object extraction based on semantic segmentation network. Long et al. proposed fully convolutional network (FCN) [13] as the pioneering work of deep learning semantic segmentation model based on full convolutional network, and subsequent algorithms have improved on this framework. Ronneberger et al. proposed U-net [14] to improve the FCN's loss of information in task practice with an encoder–decoder network structure. Then, this method was followed by several models such as DeepLab (v1/v2/v3) [15–17], multi-path refinement networks (RefineNet) [18], and pyramid scene parsing network (PSPNet) [19]. Advances in semantic segmentation network have made it possible to improve the accuracy and efficiency of marine raft aquaculture area extraction. Yueming et al. used richer convolutional features network (RCF) [20] to extract rafts through edge detection in a raft aquaculture area in Sanduao, China [21]. Shi et al. used dual-scale homogeneous convolutional neural network (DS-HCN) to extract rafts in a dual-scale full convolutional network, finding it had superior performance on marine raft aquaculture in Dalian, China [22]. Cui et al. proposed improved U-net with a pyramid upsampling and squeeze-excitation (PSE) structure (UPS-net), which captures both boundary and background information by adding PSE structures to the decoder part of U-net, with this method being effectively verified in marine raft aquaculture in eastern Lianyungang, China [23]. However, Yue's method suffers from partial edges [21]. Shi's method is mainly aimed at rafts, and the segmentation results are incomplete and suffer from the adhesion problem [22]. Cui's method has been experimentally demonstrated to be more accurate than other popular networks based on the FCN model framework. It was proposed to solve the adhesion problem of the DS-HCN method, and it is more suitable for marine aquaculture than the DS-HCN method, but it does not take advantage of the characteristics of the raft itself, and the edge of the raft is rough and incomplete [23].

It can be seen that state-of-the-art works, such as those by Chu, Fan, and Hu, mainly rely on artificial parameter adjustments and feature designs. The deep learning method for semantic segmentation avoids a large amount of manual work, but there are still poor integrity and boundary fuzzy flaws in the detection results.

This paper proposes a segmentation method, which combines a semantic segmentation network with the nonsubsampling contourlet transform (NSCT), to extract marine raft aquaculture areas and to overcome the phenomena of rough edges, adhesion, and incomplete results in the existing methods. To the best of our knowledge, this paper is the first to attempt to use a semantic segmentation network to extract marine raft aquaculture areas from SAR images.

The method is characterized by improvements in feature enhancement and model optimization on the basis of the feature analysis of marine raft aquaculture areas, as follows:

1. To address the low signal-to-noise ratio problem in SAR images, we enhanced Sentinel-1 images with the NSCT [24] to strengthen the subject contour features and directional features.
2. To capture better feature representations, we combined several modules in U-net. Multiscale convolution was used to fit the multisize characteristics of marine raft aquaculture areas, asymmetric convolution was selected to address the floating raft strip geometric features, and the attention module was adopted to focus on both spatial and channel interrelationships.

This paper is organized as follows. The first part introduces the background significance of marine raft aquaculture area extraction and the current research status, the second part analyzes the characteristics of marine raft aquaculture areas, the third part introduces the details of the method proposed in this paper, the fourth part shows the experimental results and analyzes the results, the fifth part provides the discussion, and the sixth part is the conclusion.

2. Feature Analysis of Marine Raft Aquaculture Areas

Feature analyses of the marine raft aquaculture areas provide the basis for the design of the method. A SAR image is the reflection of the target on radar beam, and the single-band echo information reflects more scattering characteristics and structural characteristics of the target. Hence, this section focuses on the scattering characteristics and structural characteristics of raft aquaculture areas.

2.1. Scattering Characteristics

Rafts are basically floating with floating balls on the surface of the water, and thus the scattering from the raft culture area consists mainly of surface scattering from the seawater and the balls, with two-sided angles and spirals scattering between them [25]. Therefore, an area where a raft exists has a different scattering intensity than areas with only seawater. Due to the presence of waves, surges, currents, and internal waves in various regions of the ocean, the backscatter characteristics of the ocean are very irregular. Furthermore, the backscatter characteristics of floating raft, influenced by the sea state, vary in different areas of the ocean. Therefore, enhancing the features of marine raft aquaculture areas in SAR images is necessary to enhance the commonalities among marine raft aquaculture areas in order to overcome the lack of floating raft features in SAR images and to mitigate background effects.

SAR images are visualized by the coherent processing of echoes from successive radar pulses, in which coherent speckle noise is unavoidable. Speckle noise exhibits a granular, black and white dotted texture on an image. Due to noise, some pixels in a homogeneous region are brighter than average, while the others are darker. Thus, the speckle effect makes a radar image of a floating raft look like a random matrix, and the magnitude values of the backscattering coefficient obey the Rayleigh distribution [26]. Raft culture increases the roughness of the sea surface, and the backscattering signal of seawater in a floating raft region is enhanced. Nevertheless, considering the influence of periodic ocean waves, backscattering coherence superposition is more prominent, resulting in more severe coherent speckle noise in SAR images [25]. As shown in Figure 2, the grayscale values of the pixels depict the amplitude values of the backscattering at each pixel, and the variability in grayscale values within the raft aquaculture area leads to the blurred edges of the area, as well as inconspicuous local features. The global features can better characterize the raft culture area.

Figure 2. Difference in gray levels of a floating raft affected by noise. (a) The raft aquaculture area is globally distinguishable. (b) The local culture area is poorly characterized. (c) Differences in grayscale values exist at various pixels within the floating raft.

2.2. Structural Characteristics

Marine raft aquaculture areas tend to extend outward from offshore areas near islands and cover wide areas with distinct structural characteristics. The structural characteristics of raft culture areas help us to distinguish them from the seawater background, which includes but not limited to the following aspects:

- Multisize characteristics:

The multisize nature of marine raft aquaculture areas is twofold. Overall, the aquaculture regions are scattered, with varying regional range sizes and inconsistent densities. Locally, the strips in the aquaculture areas are uniform in width, vary in length, and have narrow sea lanes, which vary in width between rafts. Thus, the method design needs to consider a method that can fit multisize features, and the use of a single feature sensibility field makes avoiding missing detailed information difficult.

- Strip-like geometric contour characteristics:

Floating rafts are made of ropes in series with floating balls and have distinct strip geometric characteristics in an image. The non-centric symmetry of this type of rectangle needs to be noted when using convolution to extract targets.

- Outstanding directionality:

The arrangement of floating rafts within an aquaculture area is directional, has explicit main directions, and is generally parallel to the shoreline.

Figure 3 shows the structure of the marine raft aquaculture areas. The floating rafts in areas A and B have the same strip-like geometric features and are aligned in the same direction within each zone. The size, density, and alignment direction between regions are different, i.e., zone A is more tightly packed than zone B, and the rafts are arranged horizontally in zone A and vertically in zone B.

Figure 3. (a,b) show two independent examples of marine raft aquaculture areas. Raft aquaculture areas A in (a) and B in (b) are inconsistent in range size, density, and arrangement orientation.

In summary, the scattering features cause marine raft aquaculture areas to have weak local features on SAR images and detailed features can be missed, which make it difficult to distinguish raft aquaculture areas from seawater using only scattering features. The structural characteristics indicate that when designing an approach, attention needs to be paid to the multisize and internal uniform directional features, as well as the geometric features of the floating raft strip contour. On the basis of the features of marine raft aquaculture areas, this paper proposes a segmentation method involving feature enhancement and a semantic segmentation network similar to U-net, which will be introduced in Section 3.

3. Methods

There are four main steps in the semantic segmentation method for extracting the target area: dataset construction, model construction, model training, and final testing. In addition, the accuracy of target extraction is calculated on the basis of the results of the final testing. The method for

extracting marine raft aquaculture areas in this paper adds a feature enhancement step between dataset construction and model construction. The overall process of the method in this paper is shown in Figure 4.

Figure 4. Overall flow chart.

In the dataset processing stage, this paper collected Sentinel-1 image data. After basic image processing of the data, we used the ArcGIS software to mark the image and generate binary maps called ground truth maps. Then, the images and ground truth maps were divided into training, validation, and test datasets at a 3:1:1 ratio, and data augmentation methods including mirroring, panning, and other operations on the training and validation samples were used to expand the dataset. The specific details are presented in Section 4.1.

After the construction of the dataset was completed, this paper used the NSCT to enhance the contour and orientation features of the image, and the obtained low-frequency sub-band and direction sub-bands were synthesized with the original image into a 26-channel image. This step is explained in detail in Section 3.1.

In the model construction phase of this paper, the task of detecting marine raft aquaculture areas was implemented by constructing a semantic segmentation algorithm model similar to the U-net model. The details are presented in Section 3.2.

During the model training phase, the training samples and the validation samples were input into the network model of this paper, and then the weight information was saved. The saved weights were applied to the test samples in the final testing stage to extract the floating raft region and calculate the final accuracy.

3.1. Feature Enhancement

During the feature analysis in Section 2, it was shown that the raft culture area was less distinguishable from seawater on the SAR image, but the raft culture area had significant contour and directional features. The NSCT method is well known for its capacity to highlight the main contour and directional features [27,28]. In this paper, the NSCT was used to enhance the main contour features of marine raft aquaculture areas, clarify the direction of the raft arrangement, and thus improve its distinguishability.

The NSCT is an improved method of the contour wave transform with anisotropy, multi-directionality, and translation invariance and consists of the nonsubsampled pyramid filter bank (NSPFB) and the nonsubsampled directional filter bank (NSDFB). As illustrated in Figure 5, the NSPFB acquires sub-band images with different frequencies through an iterative filter bank for multiscale decomposition of images, and the NSDFB acquires directional sub-band images with different directional divisions through a directional filter bank.

Figure 5. The two parts of the nonsubsamped contourlet transform (NSCT) for the decomposition diagram. (a) Nonsubsampled pyramid filter bank (NSPFB) decomposition diagram with the pyramid filter (y_1 is the final obtained low-frequency sub-band). (b) Nonsubsampled directional filter bank (NSDFB) decomposition diagram with the directional 2D filters (decomposition of high-frequency images).

Figure 6 shows a sample image decomposed by the NSCT. Figure 6a is the original image, Figure 6b corresponds to y_1 in Figure 5, and Figure 6c–i is the directional sub-band of the decomposed image. Figure 6b concentrates most of the energy of the original diagram and describes the main contour features well. Figure 6h is the main directional sub-band image of the original diagram, where the directionality of the floating raft arrangement can be clearly observed. It can be seen that the sub-band images obtained from the NSCT, which are decoupled from the SAR image, describe the main profile features of the floating raft and the directional features of the floating raft arrangement in the marine raft aquaculture area well and make full use of the information in the SAR images to enrich the data features. Therefore, this paper enhanced the data features with the NSCT before importing the data into the network for training. The scale parameter of the NSPFB was set to 2, and the direction parameter of the NSDFB was set to 8 according to the image size (512) of the whole scene image in the dataset.

Figure 6. Cont.

Figure 6. NSCT decomposition image results. (a) Original image. (b) Low-frequency sub-bands. (c–i) High-frequency directional sub-bands.

3.2. Fully Convolutional Networks

This section specifies the structure of the semantic segmentation network, which is similar to the U-net structure proposed in this paper for the extraction of marine raft aquaculture areas from Sentinel-1 images and includes the design of the convolution module and the integration of the attention mechanism shown in Figure 7.

The original U-net uses only a simple 3-by-3 convolution layer, which makes it difficult to fit the multisize features and regular geometric features of the raft culture area. Therefore, in the encoder stage, we introduced multiscale convolution to adapt for multisize features and asymmetric convolution in order to filter geometric features, which is shown in Figure 7b. To fully use the explicit global feature while discarding vague local features, we introduced a spatial attention mechanism at the encoder stage to calculate the spatial relationship among the pixels, and the global features were assigned to each pixel by weighting. A channel attention mechanism was adopted to direct more attention to the contour and directional features acquired through the NSCT decomposition. Figure 7c shows the attention mechanism that is used.

3.2.1. Convolution Block

As stated in Equation (1), the essence of convolution is a kind of weighted superposition. In the field of image processing, the size and weight of the convolution are designed to extract the required features from the image. The extracted multiple features constitute a multi-dimensional feature space, where inter-class variance is expected to be enhanced and intra-class difference is expected to be suppressed. In a fully convolutional network, the image is mapped to a high-dimensional feature space by means of convolutional modules, and the weights of the convolution are learned through data training to avoid the uncertainty of artificial design.

$$H = F * G$$

which can be written as

$$H(i, j) = \sum_m \sum_n F(m, n) G(i - m, j - n) \quad (1)$$

Figure 7. The model for the marine raft aquaculture area extraction task. (a) The overall model structure; (b) the Con_block with an asymmetric and multiscale structure; (c) the structure of the Attention_block with channel attention and spatial attention.

Therefore, in this paper, a full convolutional network was designed for the marine raft aquaculture area extraction task. It used multiscale convolution to extract multisize characteristics and asymmetric convolution to extract the strip-like geometric characteristics of marine raft aquaculture areas.

- **Multiscale Convolution**

Since the marine raft aquaculture areas vary in size and the spatial structures of the areas consist of large strip rafts and narrow sea lanes, we proposed the extraction of the features by a multiscale convolutional kernel, which is an appropriate choice. On the one hand, multiscale convolution extracts the information of the large-scale strip rafts and the detailed information of the narrow sea lanes. On the other hand, multiscale convolution can also capture features effectively, regardless of the size differences among the areas. When convolution kernels of different sizes, such as 3×3 , 5×5 , or 7×7 , are applied simultaneously to extract feature maps, the computational complexity of the model increases. Inspired by the GoogLeNet architecture, we designed a multiscale convolutional kernel, as shown in Figure 8a. Due to the computational characteristics

of convolution, the computational effects of two 3×3 convolution kernels are equivalent to that of a 5×5 convolution kernel, and the computational effects of three 3×3 convolution kernels are equivalent to that of a 7×7 convolution kernel [29–32]. Therefore, in this paper, the feature map fusion of multiscale convolutional kernels was achieved through series and parallel convolution kernels, which led to the extraction of features. Then, in the basic unit of each encoder, 3/5/7, three-scale receptive field information was obtained through three 3×3 convolution kernels.

Figure 8. Multiscale convolution kernel was achieved through series and parallel convolution kernels. (a) A schematic diagram of the multiscale convolutional structure advanced in this paper; (a) is equivalent to (b).

- Asymmetric convolution

The sensory field of common convolution is a rectangle with equal length and width, and thus it is difficult to capture the shape features of the non-centric symmetrical target. In consideration of the remarkable geometric structure of strip rafts, we selected asymmetric convolution kernels of sizes 1×3 and 3×1 for additive fusion with the results extracted from the 3×3 convolutional kernels.

3.2.2. Attention Block

Although the original SAR image was enhanced by the NSCT, there was still a need to better characterize the overall global features to overcome the interference of noise and the sea state. To address this problem, the proposed method combined the channel attention and spatial attention mechanisms in a series between the convolution modules. Channel attention is designed to direct attention to channels that contain the main and directional features of the raft culture after the NSCT. The spatial attention mechanism converts overall spatial relationship into weights assigned to each point of the raft culture area to better extract the global features. Convolutional block attention module (CBAM) [33] showed that channel attention and spatial attention can be used in chains, and inspired by efficient channel attention for deep convolutional neural networks (ECA-Net) [34], we simplified the calculations for channel attention weights to make them easier and faster.

- Channel Attention

The simple 2D convolution operation focuses only on the relationship among pixels within the sensory field and ignores the dependencies between channels. Channel attention links the features of each channel in order to focus on key information, such as the primary direction of the raft culture area, more effectively. As shown in Figure 9, the feature map was globally averaged to obtain a feature map of size [batch-size, channel, 1, 1]. Then, a 1×1 convolution was used to learn the correlation between each channel. Finally, the sigmoid function was used to obtain information about the weights assigned to each channel to adjust the feature information for the next level of inflow.

Figure 9. Channel attention.

- Spatial Attention

In addition to the dependency among channels, the overall spatial relationship also has a great influence on the extraction result. As shown in Figure 10, the spatial attention module first normalized the number of channels and then learnt the higher-dimensional features under a larger sensory field through convolution, thus reducing the flow of redundant information of low-dimensional features to the lower convolution and focusing on the overall information of the target.

Figure 10. Spatial attention.

Although the semantic segmentation network proposed in this paper adopts a U-shaped structure similar to U-net, it is different from U-net. The key to the difference lies in the design of the encoder. The network proposed in this paper additively merges multi-scale convolution and asymmetric convolution at the encoder stage to form basic coding units, and connects channel attention and spatial attention in series between these basic units.

4. Experiment, Results, and Analysis

This section illustrates the experiments on the test data in the study area. Section 4.1 presents the study area and experimental data. Section 4.2 proves the interpretability and generality of the proposed method through validation experiments, and Section 4.3 verifies the superiority of the method through comparative experiments. The experiments were carried out under 64-bit Linux system, using GeForce RTX 1080.

To quantify the experimental results, we used the following metrics: intersection over union (IOU) and F1-Score (F1). IOU and F1, at present, are commonly used as accuracy indexes. IOU means the intersection ratio of the predicted image and ground truth image. The F1 is an accuracy index that considers precision and recall. The formulas to calculate the metrics are as follows:

$$\text{IOU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

4.1. Study Area and Experimental Data

Currently, there is no authoritative open dataset for marine raft aquaculture area extraction, and thus this paper collected Sentinel-1 images of the Changhai region to construct a dataset to use as a basis for research.

4.1.1. Study Area

Changhai County is located in the Yellow Sea on the east side of the Liaodong Peninsula at longitude 122°13'18" E–123°17'38" E and latitude 38°55'48" N–39°18'26" N, as shown in Figure 11. It is under the jurisdiction of Dalian City and has a land area of 142.04 square kilometers, sea area of 10,324 square kilometers, coastline of 358.9 km [35], and sea use area of 244.82 square kilometers for raft culture, which is a typical large-scale marine raft aquaculture area [36].

Figure 11. Study area.

4.1.2. Dataset

With the advantages of high coverage, free access, and stable updates, the Sentinel-1 interferometric wide swath (IW) ground range detected (GRD) data from the European Space Agency (ESA)'s Copernicus project's dual polarized C-band SAR were chosen by this study as the data source (<https://vertex.daac.asf.alaska.edu/>).

The dataset contains four Sentinel-1 images from Changhai County (September 16, September 28, October 10, and October 22), each containing both vertical–horizontal (VH) and vertical–vertical (VV) polarization data. The cross-polarized VH data were less permeable than the isotropic polarized VV data, and it can be seen from the images in Figure 12 that the marine raft aquaculture area was difficult to observe with the cross-polarized data; thus the isotropic polarization (VV) image was used to construct the dataset.

Figure 12. Sentinel-1 dual polarization imaging.

It is difficult for SAR images to avoid speckle noise, which leads to a jump in the digital number (DN) value in a homogeneous region. Although the existing SAR image noise suppression methods have significantly improved image grayscale resolution, the texture information is too smooth and loses its unique features and information after denoising [37]. In this paper, before the image data were annotated, we utilized the preprocessing operations are of histogram equalization and linear stretch. On the basis of this operation, we used ArcGIS to annotate data and generate a .shp file. Then, the vector files were converted to binary images, called ground truth maps. The size of a Sentinel-1 image is too large, and thus the original images and the labeled ground truth maps were clipped into 10038 patch pairs with sizes of 512×512 . Figure 13 illustrates the main steps of dataset construction.

Figure 13. After selecting a data source, dataset construction requires three main steps: data labeling, generating a truth map, and cropping.

The complete dataset should include training data, validation data, and test data. To better verify the validity of the methods in this paper, we selected the image from Oct 16, which was not included in the dataset, as independent test data. The data used in experiments is shown in Table 1.

Table 1. Training, validation, and test data of the experiments.

Dataset	Time	Number of Images	Size	Number of Patches
Training	9.16/9.22/10.10	3	(512, 512)	7421
Validation	10.22	1	(512, 512)	2617
Testing	10.16	1	(6217, 3644)	/

4.2. Verification Experiment

4.2.1. Ablation Experiment

To prove that the strategy used in the presented method is effective, this section uses the results of the test image obtained by the original U-net (denoted by U-net), the network that introduced the attention layer (denoted by Attention_block + U-net), and the network that modified the convolutional structure (denoted by Attention_block + Con_block + U-net) and compares them with the result obtained by the proposed method (denoted by NSCT + Attention_block + Con_block + U-net).

Figure 14 shows the prediction maps for a typical region. Table 2 shows the evaluation of the whole test map segmentation results.

Table 2. Precision, recall, F1-Score (F1), and intersection over union (IOU) scores on the test data of the ablation experiment.

Method	IOU	F1	Precision	Recall
U-net	77.2%	87.1%	93.0%	81.9%
Attention_block + U-net	78.2%	87.8%	86.4%	89.2%
Attention + Con_block + U-net	81.4%	89.7%	87.4%	92.1%
NSCT + Attention + Con_block + U-net	83.0%	90.7%	89.1%	92.3%

From the results shown in Figure 14c, we can see that the segmentation result of U-net had poor integrity and relatively inward shrinking edges with burr. The result was greatly affected by speckle noise. U-net had difficulty capturing multisize information from the raft culture area, such as the narrower gaps within the floating rafts, as shown by the false positive (FP) pixels in the red box of Figure 14c; the smaller range and density of the raft culture area caused false negative (FN) pixels, as shown by the blue box in Figure 14c. There was considerable noise at the edges of the results, as shown in the green box in Figure 14c. U-net extracted features with only two tandem 3*3 convolutions, did not fully exploit the features of the floating raft, did not pay attention to the multisize information and the geometric features of the floating raft, and was highly influenced by speckle noise. There were a large number of omissions and neglected areas of interregional sea lanes in the U-net detection results, and thus the U-net segmentation result had low IOU and F1 but high precision.

As shown in Figure 14d, the addition of Attention_block effectively reduced the number of FN pixels and smoothed the edges, while the number of FP pixels was increased. In addition, the overflow and adhesion problems at the edges were serious, and the intervals between the raft culture areas were ignored. It can be seen in Table 2 that recall was improved by 7.3%, while precision decreased. Considering that the lack of information about the area inside the marine raft aquaculture area also contributed to this phenomenon, this paper overcomes this problem by designing a tailored convolutional structure.

The convolution block design with a multiscale convolution kernel and an asymmetric convolution kernel was more compatible in addressing the structures of marine raft aquaculture areas. As shown in Figure 14e, the number of FP pixels was reduced but FN pixels were displayed in areas where the outline of the subject was not visible in the yellow box. Asymmetric convolution was used to filter the striped geometric information of the raft. Multiscale convolution was used to adapt for the different

sizes and densities of the raft culture area, as well as the different sizes of the raft and seaway. The IOU increased by 3.2% compared to Attention_block + U-net.

Figure 14. Results of the ablation experiment for a subset of the study area. (a) Test image. (b) Ground truth. (c) The results obtained by the original U-net. (d) The results obtained by the network that introduced the attention layer. (e) The results obtained by the network that modified the convolutional structure. (f) The results obtained by the proposed method.

As simply changing the network structure would not be sufficient to address the low signal-to-noise ratio due to the SAR imaging mechanism and because areas with obscure subject contours were still missed, we used the NSCT in the proposed method to enhance the features. Figure 14f shows the results obtained by the proposed approach. The result was similar to the actual distribution of the raft culture areas and had fewer FN pixels in areas where the main contour was not obvious, such as

the area in the yellow box, and even single floating rafts could be extracted. The NSCT improved the information utilization and emphasized the main contour features of the floating raft, making the directional information clearer. Therefore, the feature enhancement operation increased the difference between the raft culture region and the background and helped to distinguish similarly structured but irregularly arranged waves. As shown in Table 2, the proposed method in this paper was optimal regardless of whether it used IOU or F1 as the evaluation index.

4.2.2. Applied Experiment

To verify the generality of the method, we in this section select demonstration areas in the coastal region of Shandong for experiments. Figure 15 shows the extraction results of the proposed method in subsets of the demonstration area. Table 3 shows the evaluation of the results for the whole demonstration area along the coast of Shandong.

Figure 15. Results of the applied experiment for subsets of the study area. (a,d) Test image. (b,e) Ground truth. (c,f) The results obtained by the proposed method.

Table 3. Precision, recall, F1, and IOU scores on the demonstration areas of the applied experiment.

Method	IOU	F1	Precision	Recall
NSCT + Attention + Con_block + U-net	83.8%	91.2%	96.1%	86.8%

As shown in the orange box in Figure 15d, the raft culture area in Shandong coastal area was slightly different from Changhai. Some of the floating rafts in the inner part of the marine raft aquaculture area were wider while they were arranged more sparsely. This led to larger fluctuation and more FN pixels at the edge when extracting the whole area, as shown in the orange box in Figure 15f. However, scattering characteristics and structural features of the raft culture area as a whole did not change, and thus the method proposed in this paper was also effective in this region. As for the results, the proposed method was applicable in others areas with the same characteristic performance besides Changhai.

4.3. Comparative Experiment

In recent similar studies, UPS-net [23] was demonstrated to be more accurate than other popular networks based on the FCN and more applicable than DS-HCN [22] to extract marine raft aquaculture area. Therefore, UPS-net was chosen as the comparison method to verify the superiority of the proposed method.

Figure 16 shows the prediction map for a typical region. Table 4 shows the evaluation of the whole test map segmentation results.

Figure 16. Results of the comparison experiment for a subset of the study area. (a) Test image. (b) Ground truth. (c) The results obtained by UPS-net. (d) The results obtained by the proposed method.

Table 4. Precision, recall, F1, and IOU scores on the test data of the comparative experiment.

Method	IOU	F1	Precision	Recall
UPS-net	74.8%	85.6%	91.7%	80.3%
Ours	83.0%	90.7%	89.1%	92.3%

As shown in Figure 16, the results of UPS-net showed FN pixels in the sparsely arranged or the small raft culture area and FP pixels in narrow sea lanes, with severe edge shrinkage. In contrast, the result of the proposed method was less affected by the background, had more complete edges, and provided better discrimination of sea lanes. Table 4 shows the evaluation results of the comparative experiments, and it can be seen that compared to UPS-net, the proposed method improved recall by 12%, IOU by 8.2%, and F1 by 5.1%. UPS-net adds PSE modules to U-net to obtain more contextual information and discards some redundant information at the decoder stage. However, simply adding multiscale information without considering the scattering features, geometry, and orientation of the marine raft aquaculture area makes it difficult to ensure the integrity of the extraction results. The method in this paper used asymmetric convolution to fit the geometric information of the raft culture area while adding multiscale information fusion, and used the attention mechanism and NSCT method to make better use of the scattering features and directionality of the raft culture area, which resulted in better outcomes.

5. Discussion

The state-of-the-art works for marine raft aquaculture areas are mostly dependent on professional experience. Although the deep learning method for semantic segmentation avoids a large amount of manual work, it does not work well when directly migrated to SAR images. Marine raft aquaculture areas in SAR images exhibit large differences in grayscale values (influenced by speckle noise) and distinct structural characteristics (striped contours and directionality). In consideration of these characteristics, this paper proposes a segmentation network combined with NSCT. This combination of frequency domain priors and semantic segmentation models provides a promising idea for future research in marine raft aquaculture areas extraction from SAR images.

The optimized model in this paper is more suitable for the task of marine raft aquaculture area extraction on SAR images. Although semantic segmentation models have been widely used in optical remote sensing image target extraction, these models do not yield good results by direct migration to SAR images. As shown in Figure 14c, the original U-net could not extract the raft culture area as a whole. The inward adhesion at the edges was caused by the inherent speckle noise in SAR images, which caused great loss to the detail and structure information, and this led us to focus more on global and structural information instead of local information [38]. Thus, attention module and multiscale asymmetric convolution were introduced to capture global and structural information, respectively, leading to a 4.2% increase in IOU and a 2.6% increase in F1.

Furthermore, the results show that the combination of NSCT and semantic segmentation model is useful for obtaining better results. Studies by Yin and Wang show that the naturally trained model focuses more on low-frequency information and is poorly robust to high-frequency information [39,40]. To enhance both the low and high frequency information at the same time, the proposed method decouples the original SAR image using NSCT. This improved IOU by 1.6% and F1 by 1%.

Overall, the proposed method obtained more satisfactory results than the state-of-the-art methods. However, there are some matters requiring attention. Firstly, post-processing techniques such as conditional random field (CRF) can be used to remove noise masks and obtain integration results [41], but this was not the focus of this article. Secondly, it is worth noting that the applicability of the method is related to the image resolution. Marine raft aquaculture areas extraction using the medium resolution imagery is validated in this paper, while imagery with higher resolution should be used

for refined information extraction inside the area. Transfer learning will be a good way for the model transfer between imagery with different resolutions [42].

6. Conclusions

This paper proposes a segmentation algorithm for the marine raft aquaculture area extraction using Sentinel-1 images and is characterized by feature enhancement and an improved semantic segmentation network.

1. Feature enhancement: In response to the low signal-to-noise ratio problem in SAR images, the floating raft features were enhanced using the NSCT. The low-frequency sub-band obtained by decomposing the original SAR image with the NSCT was used to enhance the contour features, and the high-frequency sub-bands were used to supplement details with direction information.
2. Improved semantic segmentation network: Multiscale feature fusion was introduced to better recognize large rafts and small seaways with less edge adhesion. Asymmetric convolution was adopted to capture the characteristics of floating raft strip distribution by screening the geometric features. Attention module was added to improve the integrity and smoothness in view of the grayscale variance in the homogeneous region of the SAR image caused by speckle noise.

In summary, the segmentation method makes full use of the scattering and structural features, and is effective in marine raft aquaculture area extraction. It is worth mentioning that the data in this paper had not been denoised, which eliminates the tedious step of extracting targets from an SAR image and has good application prospects.

In regions with poor sea conditions, the present method still suffers from errors caused by coherent spot noise enhanced by wave cascades. Therefore, further research is needed to address the issue of wave interference.

Author Contributions: Conceptualization, Y.Z.; methodology, Y.Z.; validation, Y.Z., C.W., and Y.D.; investigation, Y.Z., J.C. (Jing Chen), and Y.J. (Yongshi Jie); writing—original draft preparation, Y.Z.; writing—review and editing, Y.Z., C.W., Y.J. (Yuan Ji), and J.C. (Jingbo Chen); funding acquisition, C.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by “Nation Key R&D Program of China under grant [YFC0821900]” and “Big data on Earth in Support of Ocean Sustainable Development Goals Research [XDA19090123]”.

Acknowledgments: We would like to thank the Copernicus program of the European Space Agency for making Sentinel-1 SAR data freely available.

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

Abbreviations

FCN	Fully convolutional network
RefineNet	Multi-path refinement networks
PSPNet	Pyramid scene parsing network
RCF	Richer convolutional features network
UPS-net	Improved U-Net with a PSE structure
DS-HCN	Dual-scale homogeneous convolutional neural network
CBAM	Convolutional block attention module
ECA-Net	Efficient channel attention for deep convolutional neural networks

References

1. FAO. *World Fisheries and Aquaculture Overview 2018*; FAO Fisheries Department: Rome, Italy, 2018.
2. Zhang, Y.; He, P.; Li, H.; Li, G.; Liu, J.; Jiao, F.; Zhang, J.; Huo, Y.; Shi, X.; Su, R.; et al. Ulva prolifera green-tide outbreaks and their environmental impact in the Yellow Sea, China. *Neurosurgery* **2019**, *6*, 825–838. [[CrossRef](#)]
3. Chen, X.; Fan, W. *Theory and Technology of Fishery Remote Sensing Applications*; Science China Press: Beijing, China, 2014.

4. Deng, G.; Wu, H.Y.; Guo, P.P.; Li, M.Z. Evolution and development trend of marine raft cultivation model in China. *Chin. Fish. Econ.* **2013**, *31*, 164–169.
5. Fan, J.; Zhao, J.; An, W.; Hu, Y. Marine Floating Raft Aquaculture Detection of GF-3 PolSAR Images Based on Collective Multikernel Fuzzy Clustering. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2019**, *99*, 1–14. [[CrossRef](#)]
6. Fan, J.; Zhang, F.; Zhao, D.; Wen, S.; Wei, B. Information extraction of marine raft aquaculture based on high-resolution satellite remote sensing SAR images. In Proceedings of the Second China Coastal Disaster Risk Analysis and Management Symposium, Hainan, China, 29 November 2014.
7. Zhu, H. Review of SAR Oceanic Remote Sensing Technology. *Mod. Radar* **2010**, *2*, 1–6.
8. Chu, J.L.; Zhao, D.Z.; Zhang, F.S.; Wei, B.Q.; Li, C.M.; Suo, A.N. Monitormethod of rafts cultivation by remote sense-A case of Changhai. *Mar. Environ. Sci.* **2008**, *27*, 35–40.
9. Fan, J.; Chu, J.; Geng, J.; Zhang, F. Floating raft aquaculture information automatic extraction based on high resolution SAR images. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 3898–3901.
10. Hu, Y.; Fan, J.; Wang, J. Target recognition of floating raft aquaculture in SAR image based on statistical region merging. In Proceedings of the 2017 Seventh International Conference on Information Science and Technology (ICIST), Da Nang, Vietnam, 16–19 April 2017; pp. 429–432.
11. Geng, J.; Fan, J.; Wang, H. Weighted fusion-based representation classifiers for marine floating raft detection of SAR images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 444–448. [[CrossRef](#)]
12. Geng, J.; Fan, J.C.; Chu, J.L.; Wang, H.Y. Research on marine floating raft aquaculture SAR image target recognition based on deep collaborative sparse coding network. *Acta Autom. Sin.* **2016**, *42*, 593–604.
13. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
14. 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; Springer: Cham, Switzerland, 2015; pp. 234–241.
15. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv* **2014**, arXiv:1412.7062.
16. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
17. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [[CrossRef](#)] [[PubMed](#)]
18. Lin, G.; Milan, A.; Shen, C.; Reid, I. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1925–1934.
19. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
20. Liu, Y.; Cheng, M.M.; Hu, X.; Wang, K.; Bai, X. Richer convolutional features for edge detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 3000–3009.
21. Yueming, L.; Xiaomei, Y.; Zhihua, W.; Chen, L. Extracting raft aquaculture areas in Sanduao from high-resolution remote sensing images using RCF. *Acta Oceanol. Sin.* **2019**, *41*, 119–130.
22. Shi, T.; Xu, Q.; Zou, Z.; Shi, Z. Automatic raft labeling for remote sensing images via dual-scale homogeneous convolutional neural network. *Remote Sens.* **2018**, *10*, 1130. [[CrossRef](#)]
23. Cui, B.; Fei, D.; Shao, G.; Lu, Y.; Chu, J. Extracting raft aquaculture areas from remote sensing images via an improved U-net with a PSE structure. *Remote Sens.* **2019**, *11*, 2053. [[CrossRef](#)]
24. Da Cunha, A.L.; Zhou, J.; Do, M.N. The nonsubsampling contourlet transform: Theory, design, and applications. *IEEE Trans. Image Process.* **2006**, *15*, 3089–3101. [[CrossRef](#)]
25. Fan, J.; Wang, D.; Zhao, J.; Song, D.; Han, M.; Jiang, D. National sea area use dynamic monitoring based on GF-3 SAR imagery. *J. Radars* **2017**, *6*, 456–472.

26. Madsen, S.N. Spectral properties of homogeneous and nonhomogeneous radar images. *IEEE Trans. Aerosp. Electron. Syst.* **1987**, *4*, 583–588. [[CrossRef](#)]
27. Geng, J. Research on Deep Learning Based Classification Method for SAR Remote Sensing Images. Ph.D. Thesis, Dalian University of Technology, Dalian, China, 2018.
28. Jiao, J.; Zhao, B.; Zhang, H. A new space image segmentation method based on the non-subsampled contourlet transform. In Proceedings of the 2010 Symposium on Photonics and Optoelectronics, Chengdu, China, 19–21 June 2010; pp. 1–5.
29. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
30. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv* **2015**, arXiv:1502.03167.
31. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
32. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv* **2016**, arXiv:1602.07261.
33. Woo, S.; Park, J.; Lee, J.Y.; So Kweon, I. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
34. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–18 June 2020; pp. 11534–11542.
35. Party History Research Office of the CPC Changhai County Committee. *Changhai Yearbook (Archive) 2017*; Liaoning Nationality Publishing House: Dandong, China, 2018; pp. 29–31.
36. Yan, J.; Wang, P.; Lin, X.; Tang, C. Analysis of the Scale Structure of Mariculture Using Sea in Changhai County Based on GIS. *Ocean Dev. Manag.* **2016**, *33*, 99–101.
37. Liu, S.; Hu, Q.; Liu, T.; Zhao, J. Overview of Research on Synthetic Aperture Radar Image Denoising Algorithm. *J. Arms Equip. Eng.* **2018**, *39*, 106–112+252.
38. Yu, W. SAR Image Segmentation Based on the Adaptive Frequency Domain Information and Deep Learning. Master's Thesis, XIDIAN University, Xi'an, China, 2014.
39. Yin, D.; Gontijo Lopes, R.; Shlens, J.; Cubuk, E.D.; Gilmer, J. A Fourier Perspective on Model Robustness in Computer Vision. *Adv. Neural Inform. Process. Syst.* **2019**, *32*, 13276–13286.
40. Wang, H.; Wu, X.; Huang, Z.; Xing, E.P. High-frequency Component Helps Explain the Generalization of Convolutional Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–18 June 2020; pp. 8684–8694.
41. Hoberg, T.; Rottensteiner, F.; Feitosa, R.Q.; Heipke, C. Conditional Random Fields for Multitemporal and Multiscale Classification of Optical Satellite Imagery. *IEEE Trans. Geoenviron. Remote Sens.* **2015**, *53*, 659–673. [[CrossRef](#)]
42. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [[CrossRef](#)]

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

© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).