Article

SRSe-Net: Super-Resolution-Based Semantic Segmentation Network for Green Tide Extraction

Binge Cui, Haoqing Zhang, Wei Jing, Huifang Liu and Jianming Cui *

College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China; cuibinge@sdust.edu.cn (B.C.); zhhq951@sdust.edu.cn (H.Z.); wei_adam@sdust.edu.cn (W.J.); huifang_liu@sdust.edu.cn (H.L.)

* Correspondence: skd990455@sdust.edu.cn

Abstract: Due to the phenomenon of mixed pixels in low-resolution remote sensing images, the green tide spectral features with low Enteromorpha coverage are not obvious. Super-resolution technology based on deep learning can supplement more detailed information for subsequent semantic segmentation tasks. In this paper, a novel green tide extraction method for MODIS images based on super-resolution and a deep semantic segmentation network was proposed. Inspired by the idea of transfer learning, a super-resolution model (i.e., WDSR) is first pre-trained with high spatial resolution GF1-WFV images, and then the representations learned in the GF1-WFV image domain are transferred to the MODIS image domain. The improvement of remote sensing image resolution enables us to better distinguish the green tide patches from the surrounding seawater. As a result, a deep semantic segmentation network (SRSe-Net) suitable for large-scale green tide information extraction is proposed. The SRSe-Net introduced the dense connection mechanism on the basis of U-Net and replaces the convolution operations with dense blocks, which effectively obtained the detailed green tide boundary information by strengthening the propagation and reusing features. In addition, the SRSe-Net reduces the pooling layer and adds a bridge module in the final stage of the encoder. The experimental results show that a SRSe-Net can obtain more accurate segmentation results with fewer network parameters.

Keywords: harmful algal blooms; super-resolution; transfer learning; remote sensing images

Citation: Cui, B.; Zhang, H.; Jing, W.; Liu, H.; Cui, J. SRSe-Net: Super-Resolution-Based Semantic Segmentation Network for Green Tide Extraction. *Remote Sens.* **2022**, *14*, 710. <https://doi.org/10.3390/rs14030710>

Academic Editors: Junsheng Li, Gongliang Yu, Chen Wang and Yao Liu

Received: 15 November 2021

Accepted: 28 January 2022

Published: 2 February 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

Green tide is an algal bloom phenomenon formed by the explosive growth and aggregation of large algae (such as *Enteromorpha*) in the ocean under specific environmental conditions [1,2]. In recent years, due to global climate change and large-scale eutrophication of seawater, green algae, which have a particularly strong natural reproductive capacity, have been frequently found as large-scale outbreaks in the eastern coastal areas of China [3]. Large-scale green tide outbreaks seriously affect the marine ecological environment, threatening coastal tourism and aquaculture. Thus, outbreaks cause massive economic losses and has serious social impacts [4,5]. A green tide monitoring method based on traditional ship sailing consumes considerable manpower and materials. Satellite remote sensing technology, which has irreplaceable advantages over traditional monitoring methods, can accurately obtain information about the Earth's surface, such as the location and distribution range of green tide outbreaks, in a timely manner. The fully automatic and large-scale monitoring of green tides is of great significance to the prevention and control of green tide disasters [6]. In recent years, real-time monitoring of green tide dynamics using satellite remote sensing technology has become the best method for studying the temporal and spatial distribution patterns of macroalgae [7–10].

At present, green tide extraction methods for remote sensing images are mainly divided into four categories: threshold segmentation methods, traditional machine learning methods, mixed pixel decomposition methods, and deep learning methods. The threshold

segmentation methods mainly used the Ratio Vegetation Index (RVI) and the Normalized Difference Vegetation Index (NDVI) [11,12]. Threshold segmentation methods only consider the spectral characteristics of images, which require considerable prior knowledge and professional intervention [13]. In general, a threshold is difficult to determine due to the influence of calibration and instrument characteristics, cloud and shadow, atmosphere, and the saturation of a high biomass area with bidirectional reflectance [14]. Traditional machine learning methods need to define the spectral and spatial characteristics of green tides in advance [15,16]. Liang et al. conducted an automated macroalgal monitoring study in the Yellow Sea and the East China Sea using the extreme learning machine [17]. Xin et al. proposed a mixed pixel decomposition method based on the NDVI to achieve fine extraction of the green tide area [18]. Pan et al. proposed the use of a spectral unmixing method to estimate the green tide coverage area of GOCI images [19].

In recent years, deep learning has achieved great success in the fields of image classification, semantic segmentation, and target detection since this was first proposed in 2006 [20]. Semantic segmentation is used to understand an image from the pixel level, and its purpose is to determine an object category for each pixel in the image [21]. The purpose of the image segmentation method [22] is to divide the image into homogeneous regions. Multi-resolution segmentation is a segmentation method based on region merging. Many deep learning methods have been applied to the field of green tide monitoring [23,24]. Javier et al. proposed a macroalgae monitoring algorithm based on deep learning technology called “ERISNet”, which was used to detect large algae on the coast of Mexico [25]. Guo et al. constructed an automatic detection procedure based on the deep convolutional U-Net model to investigate green algae’s distribution characteristics in the Yellow Sea along the coast of China [26].

In the fields of astronomy and remote sensing, acquired images may be affected by various factors, e.g., optical degradation, limited sensor capacity, and observed instability of the scene, etc. The affected image may be blurred, noisy, or have insufficient spatial and temporal resolution. Super-resolution reconstruction is an important image processing technology that aims to estimate high-resolution images from one or more low-resolution observations of the same scene. As the pioneering work of deep learning in super-resolution reconstruction, SRCNN tried to fit a nonlinear mapping from low-resolution to high-resolution images through a three-layer convolution network to realize the high-resolution reconstruction of images [27]. Recent studies showed that super-resolution and semantic segmentation can promote each other [28]. Therefore, combining a green tide extraction task on low-resolution remote sensing images with the super-resolution technology based on deep learning can effectively improve the segmentation performance on low-resolution images.

In the border area of green tide with low *Enteromorpha prolifera* cover, the existing methods are prone to misclassification due to the inconspicuously detailed information of low-resolution images. Super-resolution technology can reconstruct image details and generate high-quality high-resolution remote sensing images of high quality [27,28]. In this paper, a super-resolution model was integrated into a deep semantic segmentation network to extract green tides from low-resolution remote sensing images. First, the WDSR network was used to learn the representation from low-resolution to high-resolution in the GF1-WFV image domain. Then, the mapping learned in the GF1-WFV image domain was transferred to recover the high-frequency information in the MODIS image. Finally, a deep semantic segmentation network was trained using the detail-enhanced MODIS remote sensing images.

The main contributions of this paper are as follows:

- (1) A deep semantic segmentation network (Se-Net) based on the encode–decode structure and a dense connection mechanism was constructed. By enhancing feature reuse and preserving detailed information, the fine-grained automatic extraction of green tide information in low-resolution remote sensing images was achieved.
- (2) By integrating super-resolution technology into a Se-Net, a semantic segmentation method (SRSe-Net) was proposed to extract green tides from low-resolution remote

sensing images. With the help of transfer learning, the feature mapping between low- and high-resolution images learned in the high-resolution GF1-WFV image domain was transferred to the MODIS image domain for the first time. By increasing the spatial resolution of remote sensing data, the ability to distinguish green tide patches from surrounding seawater was further improved.

2. Materials and Methods

In this section, a green tide semantic segmentation method based on transfer learning and integrating super-resolution technology is proposed. As shown in Figure 1, the framework mainly includes three stages: the super-resolution sub-model pre-training stage, the model training stage, and the relevant evaluation test stage, which we introduce. The structure of the super-resolution sub-model pre-training stage is first demonstrated in Section 2.2, and then the model training stage is presented in Section 2.3.

Figure 1. The framework of the proposed green tide extraction method.

2.1. Se-Net

The Se-Net neural network model proposed in this section employs an encoding-decoding structure as the basic structure of the model, and the overall structure is shown in Figure 2.

First, the red and near-infrared bands of the remote sensing images are selected as the input because of the obvious spectral characteristics of algae in the red and near-infrared bands. In addition, considering the sensitivity of the *NDVI* to green tides, the *NDVI* is also used as an input, and the spliced three-band remote sensing image is used as the input of the encoder module. The *NDVI* is calculated as follows:

$$NDVI = (NIR - R) / (NIR + R) \quad (1)$$

where *NIR* and *R* are the reflection values of the near-infrared and red-light bands observed by the satellite.

Then, a new codec based on U-Net is constructed, which mainly includes the following aspects. First, a dense block is used to replace convolution. The dense block module includes a multilevel convolution, a short-distance connection, and a feature mapping connection. Its dense connection mechanism can not only resolve the problem of gradient disappearance in the network but can also fully obtain the detailed boundary information of green tide patches in seawater, improving the identification error of border areas by strengthening the propagation and reusing features in green tide information. Second, we reduce the pooling layer. The main function of the pooling operation is to increase the receptive field and extract high-level semantic information, but it will also reduce the image resolution and a large amount of low-level detail information will be lost. For the green tide extraction task, preserving the detailed information in the image is more important than increasing the receptive field. Thus, the two-layer pooling operation is reduced based on U-Net. Third, we set a bridge module. A bridge module composed of three layers of dense blocks is set between codecs. On the one hand, these layers can aggregate the feature maps obtained by the previous layer; on the other hand, the number of input feature maps can be reduced to improve the computational efficiency. Fourth, two jump connections are retained. Jump connections can not only promote gradient back propagation and speed up the training process but can also help the decoder obtain more image details and recover a cleaner image.

Figure 2. Se-Net model.

The encoder extracts the features of green tide information through multilayer convolution and pooling operations, and then extracts the context information by continuously reducing the size of the feature map and expanding the receptive field to accurately locate the location of the green tide patches. The decoder continuously restores the size of the feature map through multilayer convolution and upsampling operations, reduces the receptive field, extracts high-level semantic features such as contour, texture, and border information of the green tide patches, and integrates them with low-level detail features at the end of the encoder.

Finally, the sizes of the two convolution kernels are 3×3 and 1×1 . A convolution operation of 1 output a binary segmentation diagram. Table 1 shows the complete network structure and detailed configuration of the Se-Net.

Table 1. Se-Net model.

Model	Name/Layer	Convolution Kernel	Feature Map
Input	-	-	$32 \times 32 \times 2$
Encoding model	Conv1	3×3	$32 \times 32 \times 32$
	DenseBlock1	3×3	$32 \times 32 \times 224$
	MaxPooling1	-	$16 \times 16 \times 224$
	DenseBlock2	3×3	$16 \times 16 \times 992$
	MaxPooling2	-	$8 \times 8 \times 992$
Bridge module	DenseBlock3	3×3	$8 \times 8 \times 2016$
Decoding model	Conv+Upsampling	3×3	$16 \times 16 \times 256$
	DenseBlock4	3×3	$16 \times 16 \times 2752$
	Conv+Upsampling	2×2	$32 \times 32 \times 64$
	DenseBlock5	3×3	$32 \times 32 \times 512$
	Conv2	3×3	$32 \times 32 \times 32$
Output	Conv3	3×3	$32 \times 32 \times 2$
	Conv4	1×1	$32 \times 32 \times 1$

2.2. Pre-Training Stage

For the task of super-resolution reconstruction of remote sensing images, the goal is to learn more information from the input low-resolution remote sensing image and the feature map obtained by convolution in the network. Then, this useful information is transferred to the back end of the network as much as possible to improve the image detail information. Ensuring the continuity of information flow is very important for the reconstruction of big data such as remote sensing images. Yu proposed the WDSR (wide activation for efficient and accurate image super-resolution) super-resolution network based on deep learning in 2018 [29], which ensures the continuity of information flow with less computational overhead and realizes high-resolution image reconstruction. Excellent reconstruction results were achieved. Therefore, this research uses a WDSR reconstruction to improve the spatial resolution of MODIS remote sensing images.

The WDSR network structure, which is shown in Figure 3a, is composed of three modules: a convolution module, a residual body, and a pixel shuffle. The residual body is composed of 16 residual blocks with the same structure, and the structure of the residual block is shown in Figure 3b. Pixel reorganization is a feature map amplification method that is often used in the field of super-resolution. Its main function is to obtain a high-resolution feature map from a low-resolution feature map through convolution and multichannel reorganization. The basic idea of the WDSR method is to find an optimal mapping function F between a low-resolution image and a high-resolution image through convolution and activation functions, and then reconstruct the image through this mapping function. Here, WDSR selects the mean square error (MSE) as the loss function as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n \left(F(X^{(i)}; \theta) - Y^{(i)} \right)^2 \quad (2)$$

where $X^{(i)}$ and $Y^{(i)}$ represent the input low-resolution sample and the corresponding high-resolution sample, respectively. $F(X^{(i)}; \theta)$ represents the model output, θ is the overall parameter, n is the number of samples, and i is the i -th sample.

Figure 3. (a) Schematic diagram of the WDSR structure, and (b) schematic diagram of the residual block structure.

2.3. The Architecture of SRSe-Net

There are two main differences between the proposed SRSe-Net model and the Se-Net model. First, the image after super-resolution through pre-trained WDSR instead of the original image is transmitted to the Se-Net. Second, the segmentation maps of the Se-Net are downsampled to obtain the final green tide extraction results. The overall structure of the SRSe-Net is shown in Figure 4.

Figure 4. SRSe-Net model.

First, the red, near-infrared band, and the NDVI of the low-resolution remote sensing image are selected as the input of the super-resolution network module WDSR. The WDSR module performs super-resolution reconstruction through convolution and activation functions to generate a reconstructed image whose size is doubled.

Then, the encoder performs feature extraction on the green tide information through multilayer convolution and pooling operations. After obtaining the high-level semantic feature map, the bridging module performs feature aggregation to reduce the number of feature maps and improve calculation efficiency; the decoder and skip connections are reserved. The contour, texture, and border details of the green tide patch are continuously restored to the size of the feature map through multilayer convolution and upsampling operations.

Finally, because the maximum pooling operation can suppress the errors caused by the network model parameters, it shows obvious advantages in the extraction of the target semantic information. The encoder mainly extracts high-level semantic features, and thus the two pooling operations in the encoder both use maximum pooling, while average pooling solves the problem of restricted neighborhood size and can distinguish the more obvious border of the features. Therefore, the last part of the model increases the average

pooling based on the original two convolution operations. The use of global average pooling, however, helps to ensure the integrity of the global information, for then the model can better identify and segment the border area of the green tide patch. Nevertheless, the size of the feature map can be restored through the pooling operation to adapt to the size of the label image. Backpropagation of errors is carried out during training. Table 2 shows the complete network structure and detailed configuration of the SRSe-Net.

Table 2. Complete network structure of SRSe-Net.

Model	Name/Layer	Convolution Kernel	Feature Map
Input	-	-	$32 \times 32 \times 2$
Encoding model	Conv	3×3	$32 \times 32 \times 3$
	SkipConv	5×5	$32 \times 32 \times 3$
	ResBlock $\times 16$	3×3	$32 \times 32 \times 3$
	SkipConv	3×3	$32 \times 32 \times 3$
	Add	-	$32 \times 32 \times 3$
	Shuffle	-	$64 \times 64 \times 3$
	Conv1		$64 \times 64 \times 32$
Encoding module	DenseBlock1		$64 \times 64 \times 224$
	MaxPooling1	3×3	$32 \times 32 \times 224$
	DenseBlock2		$32 \times 32 \times 992$
	MaxPooling2		$16 \times 16 \times 992$
	DenseBlock3	3×3	$16 \times 16 \times 2016$
Bridge module	Conv+Upsampling	3×3	$32 \times 32 \times 256$
	DenseBlock4	3×3	$32 \times 32 \times 2752$
	Conv+Upsampling	2×2	$64 \times 64 \times 64$
	DenseBlock5	3×3	$64 \times 64 \times 512$
	Conv2	3×3	$64 \times 64 \times 32$
	AvgPooling	-	$32 \times 32 \times 32$
Output	Conv3	3×3	$32 \times 32 \times 2$
	Conv4	1×1	$32 \times 32 \times 1$

2.4. Accuracy Evaluation

To evaluate our model, we used four metrics including *Accuracy*, *Precision*, *Recall*, and F1-score to evaluate the experimental results. The calculations are as follows:

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (4)$$

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

$$Kappa = \frac{TP + TN}{(TN + FP) \times (TN + FN) + (FN + TP) \times (FP + TP)} \quad (6)$$

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

3. Results

3.1. Data

3.1.1. Study Area

In recent years, due to the massive eutrophication of seawater, green tides have occurred more frequently in the coastal areas of eastern China. From August 2008 to May 2012, green tides broke out in large areas in the southern Yellow Sea and northern East China Sea. In this study, we deliberately selected several representative research areas in the Yellow Sea (latitude and longitude: 32°N~37°N, 119°E~124°E.) to prove the feasibility of the proposed method.

3.1.2. Data Source

GF1-WFV satellite images and MODIS images were used as training and testing datasets, respectively. The China Gaofen-1 satellite (GF-1) was successfully launched on 26 April 2013. The satellite was equipped with two 2 m resolution panchromatic/8 m resolution multispectral cameras and four 16 m resolution multispectral cameras. A single satellite simultaneously realized the combination of a high-resolution and a large width. The 2 m high-resolution achieved an imaging width of more than 60 km, and the 16 m resolution achieved an imaging width of more than 800 km, which meets the comprehensive needs of multisource remote sensing data and multiple high resolutions. For this paper, researchers chose a 16 m resolution GF1-WFV multispectral image as the training dataset. In addition, a Moderate Resolution Imaging Spectrometer (MODIS) mounted on the Terra and Aqua satellites is an important instrument for observing global biological and physical processes in the US Earth Observation System (EOS) program. MODIS data cover a wide range of bands (a total of 36 bands, and a spectral range from 0.4 μm –14.4 μm). Two channels (one band: 620 nm–670 nm; two bands: 841 nm–876 nm) have a spatial resolution of 250 m, five channels are 500 m, and the other 29 channels are 1000 m. MODIS images with 250 m resolution are mostly used for green tide monitoring. For this reason, we chose MODIS images with a spatial resolution of 250 m as the experimental dataset.

3.1.3. Data Preprocessing

As shown in Figure 5, in the data preprocessing stage, we chose the red and near-infrared bands of the MODIS image with a spatial resolution of 250 m. To fit the sample distribution of the MODIS data during the model training process, we performed the same process for the GF1-WFV image. Two bands of red and near-infrared were selected. The cutting of the selected research area and the interpretation of MODIS images were finished. In the image interpretation process, to obtain a relatively accurate ground truth image, we used the GF1-WFV image data acquired on the same date, similar time, and similar area as the MODIS image for reference. Finally, we cropped the selected study area image and ground truth map to generate training samples, where each small block corresponded to a small truth map. At the same time, we created verification samples and test samples and used Gaussian filters to downsample high-resolution GF1-WFV images to 32 m resolution, generating low-resolution samples for super-resolution sub-model pre-training. In addition, to reduce the impact of data availability, we adopted data enhancement techniques to enrich the training data for segmentation and super-resolution processes.

3.2. Experiments and Evaluation

We compared the SRSe-Net with traditional green tide extraction methods, including the NDVI, the RVI, the Support Vector Machine algorithm (SVM) [30], the Classification And Regression Tree (CART), and the Random Forest algorithm (RF). The main idea of the SVM is finding a separation hyperplane with maximum margin and mapping training samples that cannot be linearly separated in low-dimensional space to a high-dimensional space to make them linearly separable [31]. In addition, we also compared the SRSe-Net with the SegNet [32] and U-Net [33]. Multiple test images with different distribution characteristics were selected, including the concentrated green tide areas (test images 1 and

2) and the scattered green tide areas (test images 3 and 4). The qualitative and quantitative results of green tide extraction by different methods are given.

Figure 5. Data preprocessing process.

The recognition results are shown in Figures 6–9. In these figures, (a) is the low-resolution MODIS image used for testing, (b) is the super-resolution reconstructed MODIS image, and (d), (e), (f), (g), and (h) are the segmentation results produced by several traditional green tide extraction methods: NDVI, RVI, SVM, RF, and CART, respectively. Among them, for Figures 6 and 7, through multiple experiments on the test image, the best thresholds obtained by the NDVI and RVI are 0.36 and 0.36, respectively. For Figure 8, the thresholds of the NDVI and RVI are set to 0.53 and 0.56, respectively. For Figure 9, the thresholds of the NDVI and RVI are set to 0.48 and 0.56, respectively. In this figure, (i), (j), (k), and (l) are the segmentation results of the SegNet, U-Net, the proposed Se-Net, and the SRSe-Net, respectively. In these figures, the colors green, red, white, and black, represent the TP, FN, FP, and TN pixels, respectively, in the segmentation results.

Figure 6. Qualitative results for test image 1. (a) Low-resolution MODIS image; (b) Super-resolution reconstructed image; (c) Ground truth image; (d) NDVI, the best threshold value obtained by the experiment is 0.36; (e) RVI, the best threshold value obtained by the experiment is 0.46; (f) SVM; (g) RF; (h) CART; (i) SegNet; (j) U-Net; (k) Se-Net; (l) SRSe-Net.

Figure 7. Qualitative results for test image 2. (a) Low-resolution MODIS image; (b) super-resolution reconstructed image; (c) ground truth image; (d) NDVI, the best threshold value obtained by experiment is 0.36; (e) RVI, the best threshold value obtained by experiment is 0.46; (f) SVM; (g) RF; (h) CART; (i) SegNet; (j) U-Net; (k) Se-Net; (l) SRSe-Net.

Figure 8. Qualitative results for test image 3. (a) Low-resolution MODIS image; (b) super-resolution reconstructed image; (c) ground truth image; (d) NDVI, the best threshold value obtained by the experiment is 0.36; (e) RVI, the best threshold value obtained by the experiment is 0.46; (f) SVM; (g) RF; (h) CART; (i) SegNet; (j) U-Net; (k) Se-Net; (l) SRSe-Net.

Figure 6 shows the qualitative result of test image 1. From the enlarged display of Figures 6b, 7b, 8b and 9b, it can be seen that the clarity and contrast of the green tide patches after super-resolution processing substantially improved. The boundary between the green tide patch and the surrounding sea became clearer. Compared with the Se-Net, without the integrated super-resolution method, the SRSe-Net integrated with the method produced more accurate segmentation results (red areas are considerably reduced). Among several traditional green tide extraction methods, compared with the RVI and SVM methods, the NDVI method has a relatively better effect on green tide recognition, but compared to the SRSe-Net proposed in this paper, the performance was not ideal. In the segmentation results of the RVI and SVM, the prediction error areas increased substantially. In addition, the SegNet and U-Net methods based on deep learning predicted a large area of the green tide as seawater. At the same time, the areas where seawater was predicted as the green tide (white areas) were also substantially increased, and the segmentation results were poor. The result of test image 2 is similar to that of image 1. Compared with these methods, the

SRSe-Net proposed in this paper can not only improve the resolution of the image but can also obtain better segmentation results by reconstructing finer spatial details.

Figure 9. Qualitative results for test image 4. (a) low-resolution MODIS image; (b) super-resolution reconstructed image; (c) ground truth image; (d) NDVI, the best threshold value obtained by the experiment is 0.36; (e) RVI, the best threshold value obtained by the experiment is 0.46; (f) SVM; (g) RF; (h) CART; (i) SegNet; (j) U-Net; (k) Se-Net; (l) SRSe-Net.

Figures 8 and 9 show the qualitative results of test images 3 and 4, respectively. The green tide distribution in these two test images was relatively scattered. The RF predicted part of the seawater as the green tide. The other traditional machine learning methods incorrectly predicted the green tide areas with smaller patches as seawater. The SegNet method, which is based on deep learning predicted a large area of seawater as a green tide. Compared with the SegNet, the U-Net method detected the green tide area more accurately in test images with scattered green tide distributions, but its performance was still poor compared with the SRSe-Net method. The recognition results of different test images proved that the SRSe-Net method proposed in this paper has high robustness.

Regarding the quantitative results, Tables 3–6 show the corresponding quantitative results produced by different methods in different test images. The different methods are compared in detail using a variety of indicators. It can be seen from the table that the accuracy, precision, recall, and F1-score indices of the SRSe-Net proposed in this paper are generally higher than those of the other methods in the test images.

Table 3. Quantitative results for different methods on test image 1.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score
NDVI (0.36)	92.85	88.56	74.24	0.8077
RVI (0.36)	90.26	94.59	56.92	0.7107
SVM	87.33	98.59	37.92	0.5477
CART	90.61	72.66	85.88	0.7872
RF	91.08	73.84	86.58	0.7970
SegNet	85.74	85.81	35.38	0.5011
U-Net	85.64	79.38	39.25	0.5252
Se-Net	92.99	93.46	70.29	0.8023
SRSe-Net	94.90	87.89	83.86	0.8583

Table 4. Quantitative results for different methods on test image 2.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score
NDVI (0.36)	95.75	92.20	59.16	0.7207
RVI (0.36)	94.31	98.19	39.26	0.5609
SVM	92.82	99.71	22.53	0.3675
CART	95.34	71.42	82.81	0.7669
RF	95.95	77.20	79.84	0.7850
SegNet	92.50	75.62	28.19	0.4107
U-Net	94.47	76.38	58.37	0.6617
Se-Net	94.52	97.40	41.96	0.5866
SRSe-Net	96.88	89.87	74.77	0.8163

Table 5. Quantitative results for different methods on test image 3.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score
NDVI (0.53)	90.14	98.90	30.81	0.4698
RVI (0.56)	90.04	99.43	29.95	0.4603
SVM	94.46	98.03	62.18	0.7609
CART	91.83	92.79	45.96	0.6147
RF	94.29	86.68	70.57	0.7780
SegNet	89.82	61.22	77.02	0.6822
U-Net	93.93	83.02	71.94	0.7709
Se-Net	95.32	96.09	69.88	0.8092
SRSe-Net	96.63	92.64	82.83	0.8746

Table 6. Quantitative results for different methods on test image 4.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score
NDVI (0.48)	81.67	97.64	38.11	0.5482
RVI (0.56)	81.33	98.26	36.67	0.5340
SVM	89.84	93.04	70.45	0.8018
CART	91.17	86.85	82.20	0.8446
RF	88.46	92.55	65.76	0.7689
SegNet	86.44	81.11	69.78	0.7502
U-Net	89.78	93.21	70.07	0.8000
Se-Net	92.94	96.53	78.64	0.8668
SRSe-Net	93.82	92.57	85.71	0.8901

For test image 1, it can be seen from Table 3 that the methods based on deep learning has a low accuracy rate because the prediction results of these methods contain many red (predicting the green tide as seawater) and white (predicting seawater as the green tide) areas. Because the large area of the green tide is predicted to be seawater, the recall rate is low. The recall rates of the SegNet and U-Net methods are only 35.38% and 39.25%, respectively. The traditional green tide extraction method has a relatively high accuracy, but its performance is not ideal compared with the SRSe-Net method proposed in this paper. The RF and CART methods achieve better results than SRSe-Net in terms of the recall rate, but the accuracy rate is poor because most of the seawater is identified as green tides in these methods. It is worth noting that because the SVM is more sensitive to areas with obvious green tides, or if the SVM only identifies areas with more obvious green tides, it is poorly recognized for areas with inconspicuous green tides, resulting in a higher accuracy rate, reaching 98.59%, while the recall rate is only 37.92%. The accuracy of the segmented subnet Se-Net is 0.14% higher than that of the highest NDVI method. The accuracy of the proposed SRSe-Net is 2.05% higher than that of the NDVI, which is 1.91% higher than the segmented subnet Se-Net, and the recall rate is higher. Compared with other methods, there has been a large improvement. For test image 2, as shown in Table 4, traditional

methods have a higher accuracy rate but a lower recall rate. For instance, the RVI and SVM methods have recall scores of only 39.26% and 22.53%, respectively. The method based on deep learning has the same problem, and the recall rate of the SegNet method is only 28.19%. Compared with several other methods, the accuracy of the segmented subnet Se-Net is not as accurate as the NDVI method with the highest accuracy, but the accuracy of the SRSe-Net is 1.13% higher than that of the NDVI method.

In test image 3, as shown in Table 5, in the traditional green tide extraction method, most green tide areas are misidentified as seawater, resulting in a low recall rate; in particular, the recall rates of the NDVI and RVI methods are only 30.81% and 29.95%, respectively. As shown in Figure 10, the traditional methods are more accurate in predicting the areas where green tides exist, which leads to the accuracy of the NDVI and RVI methods reaching 98.90% and 99.43%, respectively. In addition, the method based on deep learning has a relatively high accuracy, but its performance is not ideal compared with the SRSe-Net proposed in this paper. The accuracy of the proposed segmentation subnet Se-Net is 0.86% higher than that of the highest SVM method, the accuracy of the SRSe-Net is 2.17% higher than that of the SVM, and the accuracy of the SRSe-Net is 1.31% higher than that of the segmentation subnet Se-Net. In test image 4, as shown in Table 6, both the traditional method and the deep learning-based green tide extraction method have a high accuracy and precision, but the recall rate is low. Compared with other methods, the accuracy of the proposed segmentation subnet Se-Net is 1.77% higher than that of the highest CART, and the accuracy of the SRSe-Net is 2.65% higher than that of the highest CART. The accuracy rate of the SRSe-Net is 0.88% higher than that of the segmented subnet Se-Net. Therefore, the detailed results provided in Tables 3–6 also prove that the accuracy and robustness of the SRSe-Net method proposed in this paper are considerably higher than those of the other methods and that this method more accurate green tide segmentation results can be obtained.

Figure 10. Prediction results of SRSe-Net on test images 1–8. (a) Prediction results of SRSe-Net on test image 1; (b) Prediction results of SRSe-Net on test image 2; (c) Prediction results of SRSe-Net on test image 3; (d) Prediction results of SRSe-Net on test image 4; (e) Prediction results of SRSe-Net on test image 5; (f) Prediction results of SRSe-Net on test image 6; (g) Prediction results of SRSe-Net on test image 7; (h) Prediction results of SRSe-Net on test image 8.

To further test the robustness of the proposed SRSe-Net, we applied the model to more image data. Figure 10 shows the recognition results of the SRSe-Net in eight MODIS test images, of which the last four images have not appeared previously. The quantitative results of the test images can be seen in Table 7. On all of the test images, the SRSe-Net has the highest accuracy of 96.88%, the lowest of 90.73%, the average of 93.13%, the highest

F1-score of 0.8942, the lowest of 0.8023, and the average of 0.8558. The above experimental results show that the prediction ability of the SRSe-Net is excellent and reliable.

Table 7. Quantitative results of SRSe-Net on test images 1–8.

Images	Accuracy (%)	Precision (%)	Recall (%)	F1-Score
Image 1	92.99	93.46	70.29	0.8023
Image 2	96.88	89.87	74.77	0.8163
Image 3	96.63	92.64	82.83	0.8746
Image 4	93.82	92.57	85.71	0.8901
Image 5	90.73	89.22	89.63	0.8942
Image 6	90.94	85.17	89.83	0.8744
Image 7	90.99	88.57	75.30	0.8140
Image 8	92.07	92.72	83.88	0.8808

4. Discussion

Compared with the U-Net and SegNet, the Se-Net introduced dense blocks and reduced the amount of downsampling, which significantly reduced the number of model parameters and improved the performance in the green tide border area. By super-resolution processing of MODIS images, the SRSe-Net made the task of green tide extraction easier, and the final extracted green tide coverage was more accurate. This also proved that the mapping relationship from low- to high-resolution learned from one remote sensing image domain can be effectively transferred to another image domain, even if their resolutions are quite different.

Many spheres are affected by the image quality and resolution, such as shipwreck detection using satellite remote sensing. Super-resolution technology can improve the clarity of remote sensing images. Theoretically, the method proposed by Baeye et al. [34] can be further improved using super-resolution technology to extract more precise suspended particulate materials for more reliable wreck detection results. Super-resolution techniques can also be applied to images that are obtained by other measurement methods, such as sonar images. The mapping learned by using high-resolution sonar images can be transferred to low-resolution sonar images to improve image quality, which can make the underwater information obtained from sonar images more comprehensive and sufficient. Thus, the super-resolution technology may help Janowski et al. [35] achieve fine mapping of benthic habitats.

Although the SRSe-Net obtained better results in the test images, there are still some missed extractions in the areas with low green tide coverage. This may be because the green tide is located underwater or the green tide patch is small. Due to the strong absorption of sunlight by seawater, the spectral characteristics of green tides are not obvious.

In the next work, we will try to obtain high-resolution satellite data (e.g., GF1-WFV and Landsat8) and MODIS data in the same area and at the same time. After registering multi-source remote sensing images, we can build super-resolution models at higher multiples. In addition, using the interpretation results of high-resolution images as ground truth maps, we can build a deep learning-based inversion model for green tide coverage in low-resolution MODIS images.

5. Conclusions

In this paper, we presented a novel semantic segmentation method for green tide extraction from low-resolution remote sensing images. The proposed SRSe-Net has three main advantages. The first advantage is that the threshold selection step in the traditional green tide extraction method is omitted and the green tide information can be automatically extracted from the remote sensing image without human intervention. The second advantage is that image super-resolution reconstruction technology is integrated into the semantic segmentation network as a super-resolution subnet. The representations learned from high-resolution GF1-WFV images are used to reconstruct low-resolution MODIS

images, which improves the final segmentation performance by reconstructing finer spatial details. The third advantage is that skip connections are used to transfer the detailed information of the corresponding layer from the encoder to the decoder. By strengthening the propagation and reusing features, we can obtain a higher segmentation accuracy and more accurate segmentation results. To prove the effectiveness of the proposed method, we constructed two datasets, GF1-WFV and MODIS, with different resolutions in the source and target domains. A large number of experimental results show that the performance of the proposed method is substantially better than that of other methods.

Author Contributions: Conceptualization, B.C.; methodology, B.C. and H.Z.; software, H.Z.; validation, W.J., H.L. and J.C.; investigation, B.C.; resources, J.C.; writing—original draft preparation, H.Z.; writing—review and editing, B.C., H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Shandong Province Natural Science Foundation of China under Grant ZR2020MD096 and Grant ZR2020MD099.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors would like to thank all reviewers and editors for their comments on this paper.

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

References

1. Liu, D.Y.; Keesing, J.K.; He, P.; Wan, Z.L.; Shi, Y.J.; Wang, Y.J. The world's largest macroalgal bloom in the Yellow Sea, China: Formation and implications. *Estuar. Coast. Shelf Sci.* **2013**, *129*, 2–10. [\[CrossRef\]](#)
2. Fan, S.L.; Fu, M.Z.; Li, Y.; Wang, Z.L.; Fang, S.; Jiang, M.J.; Wang, H.P.; Sun, P.; Qu, P. Origin and development of Huanghai (Yellow) Sea green-tides in 2009 and 2010. *Acta Oceanol. Sin.* **2012**, *34*, 187–194.
3. Qiu, Y.H.; Lu, J.B. Advances in the monitoring of Enteromorpha prolifera using remote sensing. *Ecol. Sin.* **2015**, *35*, 4977–4985.
4. Xing, Q.; An, D.; Zheng, X.; Wei, Z.; Chen, J. Monitoring seaweed aquaculture in the Yellow Sea with multiple sensors for managing the disaster of macroalgal blooms. *Remote Sens. Environ.* **2019**, *231*, 111279. [\[CrossRef\]](#)
5. Xiao, Y.F.; Zhang, J.; Cui, T.W.; Gong, J.L.; Liu, R.J.; Chen, X.Y.; Liang, X.J. Remote sensing estimation of the biomass of floating Ulva prolifera and analysis of the main factors driving the interannual variability of the biomass in the Yellow Sea. *Abbreviated* **2019**, *140*, 330–340. [\[CrossRef\]](#)
6. Chen, Y.; Sun, D.Y.; Zhang, H.L.; Wang, S.Q.; Qiu, Z.F.; He, Y.J. Remote-Sensing Monitoring of Green Tide and Its Drifting Trajectories in Yellow Sea Based on Observation Data of Geostationary Ocean Color Imager. *Acta Opt. Sin.* **2020**, *40*, 0301001. [\[CrossRef\]](#)
7. Qiu, Z.F.; Li, Z.X.; Bilal, M.; Wang, S.Q.; Sun, D.Y.; Chen, Y.L. Automatic method to monitor floating macroalgae blooms based on multilayer perceptron: Case study of Yellow Sea using GOCI images. *Opt. Express* **2018**, *26*, 26810–26829. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Lu, T.; Lu, Y.C.; Hu, L.B.; Jiao, J.N.; Zhang, M.W.; Liu, Y.X. Uncertainty in the optical remote estimation of the biomass of Ulva prolifera macroalgae using MODIS imagery in the Yellow Sea. *Opt. Express* **2019**, *27*, 18620–18627. [\[CrossRef\]](#)
9. Xiao, Y.F.; Zhang, J.; Cui, T.W. High-precision extraction of nearshore green tides using satellite remote sensing data of the Yellow Sea, China. *Int. J. Remote Sens.* **2017**, *38*, 1626–1641. [\[CrossRef\]](#)
10. Cui, T.W.; Liang, X.J.; Gong, J.L.; Tong, C.; Xiao, Y.F.; Liu, R.J.; Zhang, X.; Zhang, J. Assessing and refining the satellite-derived massive green macro-algal coverage in the Yellow Sea with high resolution images. *ISPRS J. Photogramm. Remote Sens.* **2018**, *144*, 315–324. [\[CrossRef\]](#)
11. Shi, W.; Wang, M.H. Green macroalgae blooms in the Yellow Sea during the spring and summer of 2008. *J. Geophys. Res.* **2009**, *114*, C12. [\[CrossRef\]](#)
12. Jordan, C.F. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* **1969**, *50*, 663–666. [\[CrossRef\]](#)
13. Guo, N. Vegetation index and its research progress. *Arid. Meteorol.* **2003**, *21*, 71–75.
14. Luo, Y.; Xu, J.H.; Yue, W.Z. Research on vegetation indices based on the remote sensing images. *Ecol. Sci.* **2005**, *24*, 75–79.
15. Chen, Y.; Dai, J.F. Extraction methods of cyanobacteria bloom in Lake Taihu based on RS data. *Lake Sci.* **2008**, *20*, 179–183.
16. Wang, R.; Wang, C.Y.; Li, J.H. An intelligent divisional green tide detection of adaptive threshold for GF-1 image based on data mining. *Acta Oceanol. Sin.* **2019**, *41*, 131–144.
17. Liang, X.J.; Qin, P.; Xiao, Y.F. Automatic Remote Sensing Detection of Floating Macroalgae in the Yellow and East China Seas Using Extreme Learning Machine. *J. Coast. Res.* **2019**, *90*, 272–281. [\[CrossRef\]](#)

18. Xin, L.; Huang, J.; Liu, R.J.; Zhong, S.; Xiao, Y.F.; Wang, N.; Cui, T.W. Effects of regional pollution on aerosol optical properties and radiative forcing in background area by ground-based and satellite remote sensing observation. *Acta Laser Biol. Sin.* **2015**, *6*, 585–589.
19. Pan, B.; Shi, Z.W.; An, Z.Y.; Jiang, Z.G.; Ma, Y. A Novel Spectral-Unmixing-Based Green Algae Area Estimation Method for GOCI Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 437–449. [[CrossRef](#)]
20. Hinton, G.; Osindero, S.; Teh, Y.W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [[CrossRef](#)]
21. Vo, D.M.; Lee, S.W. Semantic image segmentation using fully convolutional neural networks with multi-scale images and multi-scale dilated convolutions. *Multimed. Tools Appl.* **2018**, *77*, 18689–18707. [[CrossRef](#)]
22. Lu, L.; Jiang, H.; Zhang, H.J. A robust audio classification and segmentation method. In Proceedings of the Ninth ACM International Conference on Multimedia, Ottawa, ON, Canada, 30 September–5 October 2001; pp. 203–211.
23. Yang, S.; Chen, L.F.; Shi, Y.; Mao, Y.M. Semantic segmentation of blue-green algae based on deep generative adversarial net. *J. Comput. Appl.* **2018**, *38*, 1554–1561.
24. Valentini, N.; Balouin, Y. Assessment of a Smartphone-Based Camera System for Coastal Image Segmentation and Sargassum monitoring. *J. Mar. Sci. Eng.* **2019**, *8*, 23. [[CrossRef](#)]
25. Arellano-Verdejo, J.; Lazcano-Hernandez, H.E.; Cabanillas-Terán, N. ERISNet: Deep neural network for Sargassum detection along the coastline of the Mexican Caribbean. *PeerJ* **2019**, *7*, e6842. [[CrossRef](#)]
26. Guo, Y.; Gao, L.; Li, X.F. Distribution Characteristics of Green Algae in Yellow Sea Using an Deep Learning Automatic Detection Procedure. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 3499–3501.
27. Chao, D.; Chen, C.; He, K.; Tang, X. Learning a Deep Convolutional Network for Image Super-Resolution. *Lect. Notes Comput. Sci.* **2014**, *8692*, 184–199.
28. Tang, Z.J.; Pan, B.; Liu, E.H.; Xu, X.; Shi, T.Y.; Shi, Z.W. SRDA-Net: Super-Resolution Domain Adaptation Networks for Semantic Segmentation. *arXiv* **2020**, arXiv:2005.06382.
29. Yu, J.H.; Fan, Y.C.; Yang, J.C.; Xu, N.; Wang, Z.W.; Wang, X.C.; Huang, T. Wide Activation for Efficient and Accurate Image Super-Resolution. *arXiv* **2018**, arXiv:1808.08718.
30. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
31. Cui, B.G.; Zhong, L.W.; Yin, B. Hyperspectral Image Classification Based on Multiple Kernel Mutual Learning. *Infrared Phys. Technol.* **2019**, *99*, 113–122. [[CrossRef](#)]
32. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)]
33. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
34. Baeye, M.; Quinn, R.; Deleu, R.; Fettweis, M. Detection of shipwrecks in ocean colour satellite imagery. *J. Archaeol. Sci.* **2016**, *66*, 1–6. [[CrossRef](#)]
35. Janowski, L.; Wroblewski, R.; Dworniczak, J.; Kolakowski, M.; Rogowska, K.; Wojcik, M.; Gajewski, J. Offshore benthic habitat mapping based on object-based image analysis and geomorphometric approach. A case study from the Slupsk Bank, Southern Baltic Sea. *Sci. Total Environ.* **2021**, *801*, 149712. [[CrossRef](#)] [[PubMed](#)]