



Article

HY1C/D-CZI *Noctiluca scintillans* Bloom Recognition Network Based on Hybrid Convolution and Self-Attention

Hanlin Cui ¹, Shuguo Chen ^{1,2,3,*}, Lianbo Hu ¹, Junwei Wang ¹, Haobin Cai ², Chaofei Ma ³, Jianqiang Liu ³ and Bin Zou ³

¹ College of Marine Technology, Ocean University of China, Qingdao 266100, China

² Sanya Ocean Institute, Ocean University of China, Sanya 572024, China

³ National Satellite Ocean Application Service, Ministry of Natural Resources of the People's Republic of China, Beijing 100081, China

* Correspondence: chenshuguo@ouc.edu.cn

Abstract: Accurate *Noctiluca scintillans* bloom (NSB) recognition from space is of great significance for marine ecological monitoring and underwater target detection. However, most existing NSB recognition models require expert visual interpretation or manual adjustment of model thresholds, which limits model application in operational NSB monitoring. To address these problems, we developed a *Noctiluca scintillans* Bloom Recognition Network (NSBRNet) incorporating an Inception Conv Block (ICB) and a Swin Attention Block (SAB) based on the latest deep learning technology, where ICB uses convolution to extract channel and local detail features, and SAB uses self-attention to extract global spatial features. The model was applied to Coastal Zone Imager (CZI) data onboard Chinese ocean color satellites (HY1C/D). The results show that NSBRNet can automatically identify NSB using CZI data. Compared with other common semantic segmentation models, NSBRNet showed better performance with a precision of 92.22%, recall of 88.20%, F1-score of 90.10%, and IOU of 82.18%.

Keywords: *Noctiluca scintillans* blooms; remote sensing; deep learning; HY-1C/D; CZI



Citation: Cui, H.; Chen, S.; Hu, L.; Wang, J.; Cai, H.; Ma, C.; Liu, J.; Zou, B. HY1C/D-CZI *Noctiluca scintillans* Bloom Recognition Network Based on Hybrid Convolution and Self-Attention. *Remote Sens.* **2023**, *15*, 1757. <https://doi.org/10.3390/rs15071757>

Academic Editor: SeungHyun Son

Received: 2 March 2023

Revised: 19 March 2023

Accepted: 21 March 2023

Published: 24 March 2023



Copyright: © 2023 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

Noctiluca scintillans is widely distributed in global nearshore waters [1], occurring frequently worldwide in recent years [2], especially along the coast of China [3,4]. Although *Noctiluca scintillans* blooms (NSBs) are non-toxic red tides, explosive proliferation, and aggregation of *Noctiluca scintillans* cells makes the ocean surface mucilaginous or gelatinous [2], seriously impeding the exchange of sea air in the surface layer of the ocean and affecting the marine ecosystem [5,6]. *Noctiluca scintillans* emits blue light when mechanically disturbed by waves, ships, underwater vehicles, etc. [7,8], which can be applied for surface and underwater target detection at night [9,10]. Therefore, an accurate and efficient automated recognition method for NSB detection is of great significance for both marine ecological studies and underwater target detection.

Currently, NSB can be identified by field experiments of remote sensing. In the German Bight, Schaumann et al. [11] established a link between water column eutrophication and the abundance of *Noctiluca scintillans* cells. Uhlig et al. [12] found a pattern of seasonal periodic outbreaks of NSB through long-term observations. In the East China Sea, Tseng et al. [13] conducted high-frequency sampling at fixed stations and found that the area beyond the estuary of the Yangtze River was the main active area of the NSB. However, most of the above field-based studies are limited to a shorter period and a smaller area, which leads to findings that are constrained in time and space. Satellite remote sensing has been widely used in various red tide recognition tasks with the advantages of a fixed revisit period, low cost, and high spatial and temporal coverage. At present, a large number of studies have proposed various threshold-based band ratio or baseline subtraction

methods based on the analysis of the spectral characteristics of red tide. For example, Tang et al. [14] demonstrated the feasibility of the normalized difference vegetation index in red tide recognition from remote sensing data. Hu et al. [15] used fluorescence line height to detect and track a red tide outbreak off the coast of southwest Florida using MODIS data. Ahn et al. [16] designed the red tide index (RI) based on red tide spectral features and applied it to SeaWiFS data to identify red tides off the coasts of Korea and China. Takahashi et al. [17] further improved the RI based on spectral features and applied it to Sentinel-2 MSI data to identify red tides in lakes [18]. Based on the unique absorption and scattering properties of NSB, Qi et al. [19,20] determined the NSBs using the ratio of multiband differences and analyzed the impacts of environmental factors on NSB distribution. Unlike the direct recognition of NSBs from spectral features, Dwived et al. [21] used sea surface temperature and chlorophyll to identify NSBs in the Arabian Sea. By analyzing the spectral response characteristics of red tide, Liu et al. [22] designed a red tide recognition method for GF-1 WFV data and validated it in a red tide event. To expand the differences between red tide and non-red tide water bodies, Liu et al. [23] chose to transfer remote sensing images from RGB color space to CIE1931 color space. However, there are many difficulties when applying index-based algorithms to identify or monitor NSBs. First, identifying small *Noctiluca scintillans* patches in satellite remote sensing data is challenging [19]. Second, passive remote sensing satellites are susceptible to interference from various influencing factors such as clouds, aerosols, and observation geometry [24]. Finally, almost all NSB recognition methods require manually adjusted thresholds or expert visual examination, making them difficult to operate automatically.

With powerful feature extraction and nonlinear approximation capabilities, deep learning has achieved excellent results in classification, detection, fusion, and segmentation using remote sensing data. In 2015, a fully convolutional network (FCN) was first applied to the semantic segmentation task [25]. Then, Ronneberger et al. [26] developed UNet, a semantic segmentation model using an encoder-decoder structure, where the encoder is used to extract semantic information from the input data and the decoder is responsible for fusing the semantic information with the underlying features and finely decoding the segmentation results. A jump connection is set between the encoder and the decoder to preserve the underlying details and reduce the feature loss due to downsampling. With the innovative work of the encoder-decoder, UNet achieves better results when applied in remote sensing with fewer training samples. The subsequently developed SegNet [27] and DeepLab V3+ [28] adopting the encoder-decoder structure were applied in red tide recognition. Lee et al. [29] developed a neural network to identify and monitor red tides using high-resolution satellite data. Kim et al. [30] verified the possibility of using UNet for automatic pixel-based red tide detection. Later, Zhao et al. [24] added a channel attention model to UNet for better identification of red tide using HY-1D and GF-1 data.

Although the deep learning models mentioned above can recognize NSB using remote sensing images, the simple stacked convolution operation cannot fully utilize the information from remote sensing data. To efficiently utilize computational resources and extract diverse features from remote sensing data, various deep learning-based approaches have emerged. In 2014, the Google team first proposed the concept of Inception in GoogLeNet [31], and its core idea is to extract multiscale information by combining different convolutional parallel structures into a sparse network structure and finally, perform feature fusion to obtain a better image representation. Later, Google proposed Inception V2–V4 [32–34] and Xception [35] to reduce the computational effort by using parallel convolution while maintaining the multiscale feature extraction capability. Liu et al. combined Inception with dilated convolution to further exploit the potential of Inception [36]. In addition to multiscale feature extraction, other studies tend to model global correlations within data using self-attention learning. Lin et al. first introduced the concept of self-attention in 2017 [37], which was designed to enable networks to focus on globally important features and ignore irrelevant distracting information. Recent studies have shown that self-attention is also applicable to computer vision tasks. Vision in Transformer (ViT)

proposed in [38] first applied a self-attention-based Transformer to image classification with better results than CNN by exploring the connections between global data through self-attention learning. Since self-attention can build long-range dependencies in data and learn high-quality intermediate features, it was applied to DANet [39] and CCNet [40] to extend CNN capacities. On the other hand, CvT [41], CoAtNet [42], UniFormer [43], and iFormer [44], pair self-attention with convolution. Later, SETR [45] used ViT exclusively as the encoder, PVT [46] used ViT encoder while giving the model the ability to extract multiscale features by setting up a pyramid structure, and these improvements further increased the computational cost of the model. To reduce the huge computational effort required by self-attention, Liu et al. [47] designed the Swin Transformer in 2021 by introducing a CNN-like hierarchical construction method that combines the windowing of convolution and the global nature of self-attention to achieve the best performance in a variety of vision tasks while having low computational complexity. The success of ViT in vision tasks shows the potential of self-attention in remote sensing, inspiring a series of self-attention-based small-data networks, such as UTNet [48], Transformer UNet [49], Trans UNet [50], and Swin UNet [51], which have been further applied in remote sensing change detection [52–54] and semantic segmentation [55–58].

Inspired by the above studies, to automatically recognize NSBs from remote sensing data and improve the recognition accuracy, this manuscript proposes a *Noctiluca scintillans* Bloom Recognition Network (NSBRNet) incorporating convolution and self-attention to achieve automatic recognition of NSBs using high spatial resolution satellite remote sensing data. The rest of this manuscript is organized as follows. The second part describes the Coastal Zone Imager (CZI) remote sensing data and the processing methods. The third part presents the details of NSBRNet. The fourth part examines the recognition accuracy of the model and conducts comparison and ablation experiments to compare with other deep learning models and to examine the performance of ICB, SAB, and NSBI.

2. Materials and Methods

2.1. CZI Data and NSB Event Information

CZI Level-1B data were acquired from the National Satellite Ocean Application Service (NSOAS). The CZI sensor was carried on the HY-1C satellite launched in 2018 and the HY-1D satellite launched in 2020. The CZI sensor has four bands with a spectral range of 460–825 nm, swath of 950 km, and spatial resolution of 50 m. Onboard HY-1C and HY-1D, CZI can make observations with a revisit period of 1 day at any location along the China coast. With a wide swath and high spatial resolution, CZI sensors are widely used in various ocean monitoring missions. Figure 1 shows an example of NSB in the East China Sea acquired by CZI on 17 August 2020. The NSB event information used in this study was obtained from the reports released by NSOAS. CZI data on cloud-free days were searched to match the NSB events. Finally, six CZI images were matched with reported NSB events and are summarized in Table 1.

Table 1. Information on NSBs in CZI.

Date	Region	Longitude	Latitude
17 August 2020	East China Sea	123°36′–125°32′	31°58′–33°18′
17 August 2020	East China Sea	124°92′–135°45′	32°63′–34°73′
14 February 2021	Beibu Gulf	107°71′–109°38′	19°21′–21°25′
13 March 2022	Dapeng Bay	108°84′–118°15′	19°49′–21°45′
10 April 2022	Yangjiang	111°45′–120°83′	20°08′–22°06′
10 April 2022	Dapeng Bay	112°05′–121°62′	22°90′–24°90′

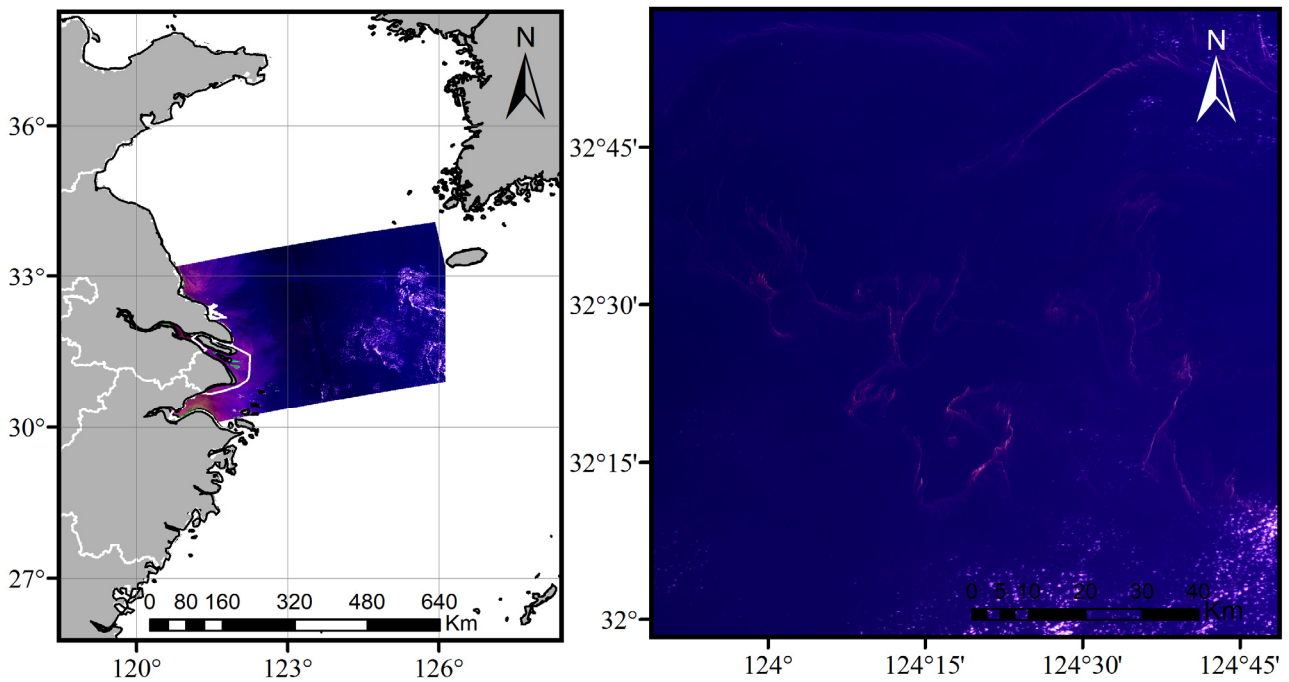


Figure 1. CZI FRGB image with NSB acquired on 17 August 2020.

CZI Level-1B data were processed to Rayleigh corrected reflectance (R_{rc}) [59]:

$$R_{rc} = R_t - R_r \quad (1)$$

$$R_t = \pi L_t / (F_0 \cos \theta_0) \quad (2)$$

where R_t denotes top-of-atmosphere reflectance, R_r denotes Rayleigh reflectance, L_t denotes total top-of-atmosphere irradiance, F_0 denotes extraterrestrial solar irradiance, and θ_0 denotes solar zenith angle. The Rayleigh corrected CZI reflectance data were further subjected to geometric correction, *Noctiluca scintillans* bloom annotation, data cropping, data partitioning, sample equalization, and data augmentation. The specific data processing flow chart is shown in Figure 2:

2.2. *Noctiluca scintillans* Bloom Index

Influenced by wave and tidal motion, *Noctiluca scintillans* patches are usually stretched into thin strips on the ocean surface (Figure 3a). Unlike the general semantic segmentation task, the spectral features of the *Noctiluca scintillans* patches vary with the abundance of *Noctiluca scintillans* cells per unit area. As shown in Figure 3b, high spectral reflectance in the red band is observed in intense *Noctiluca scintillans* patches (points 1 and 2), while relatively low reflectance is observed in sparse patches (points 3 and 4).

To enhance the difference between the NSB and the background waters, similar to the Floating Algae Index (FAI) proposed by Hu [60], the *Noctiluca scintillans* Bloom Index (NSBI) was designed in this study using the green and near-infrared bands as the baseline to measure the reflectance peaks of *Noctiluca scintillans* in the red band. The NSBI is calculated as follows:

$$NSBI = R_{rcR} - R'rc_R \quad (3)$$

$$R'rc_R = R_{rcNIR} + (R_{rcG} - R_{rcNIR}) \times (\lambda_{NIR} - \lambda_R) / (\lambda_{NIR} - \lambda_G) \quad (4)$$

where the subscripts *NIR*, *R*, and *G* are wavelengths in the near-infrared, red, and green regions, respectively. For the CZI data, $\lambda_G = 560$ nm, $\lambda_R = 650$ nm, and $\lambda_{NIR} = 825$ nm. The Rayleigh-corrected reflectance of four typical targets on the sea surface is shown in Figure 4a, and the corresponding NSBI is shown in Figure 4b. The NSBI values of water

and clouds are close to zero with low variation, while the NSBI values of *Noctiluca scintillans* are greater than zero. Macroalgae show a negative NSBI value due to the strong reflection peak in the near-infrared band. Clearly, the NSBI can be used to differentiate NSBs from other ocean targets, such as water, clouds, and macroalgae.

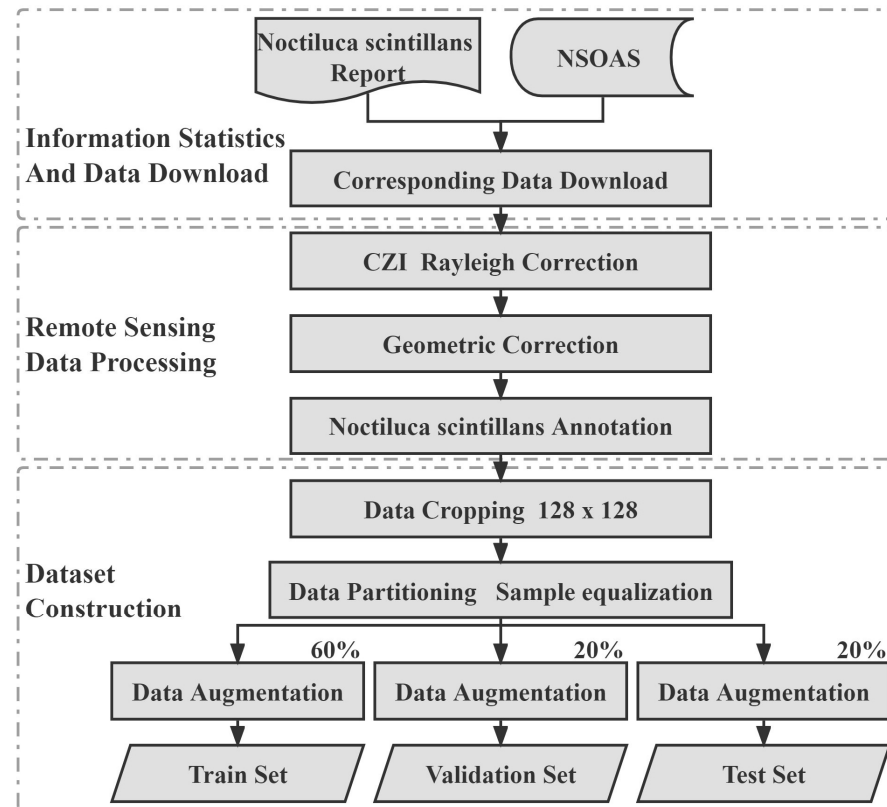


Figure 2. Flow chart of CZI data processing.

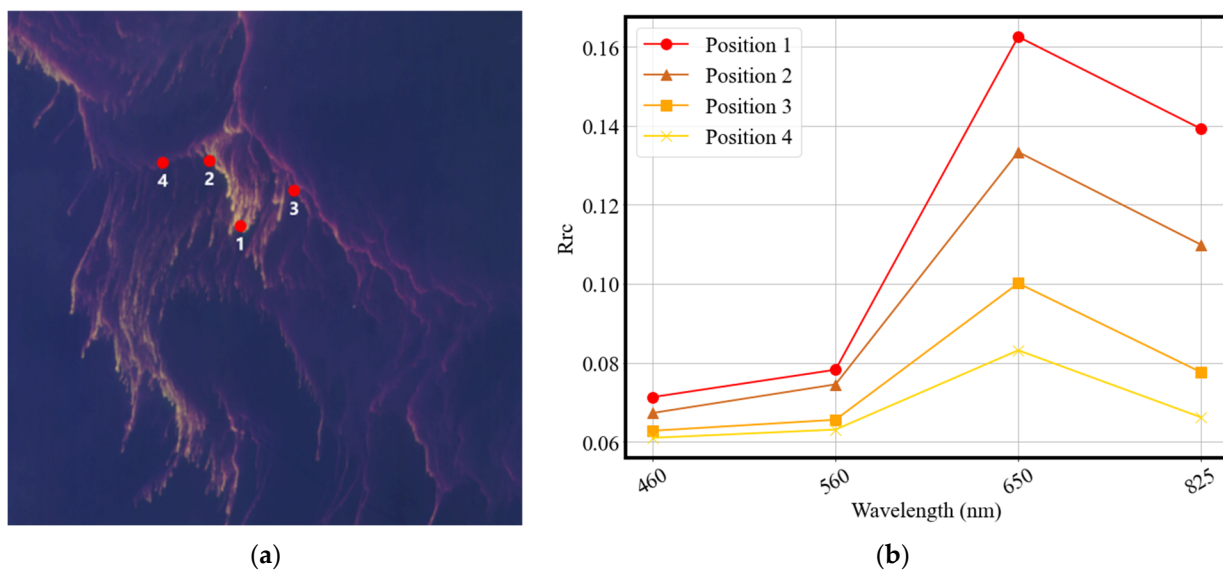


Figure 3. Spectral reflectance of NSB at different locations in the CZI image. (a) Locations of four points with different abundances; (b) Spectral reflectance of NSB at the four points.

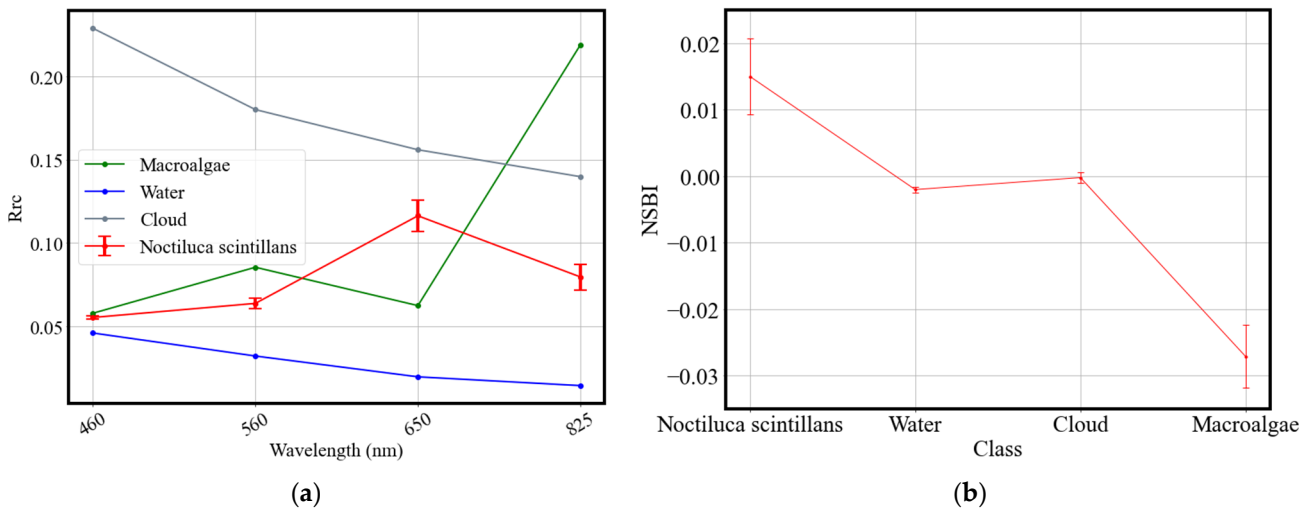


Figure 4. (a) Spectral reflectance and (b) the corresponding NSBI values of four typical targets on the sea surface.

2.3. Annotation Method

Similar to Qi et al. [19], the water color anomaly region was first identified by visual interpretation, and then the R_{rc} of the target pixel was subtracted from the clean water pixel within 3 × 3 km near the target pixel to obtain ΔR_{rc} to minimize the influences of thin clouds or aerosols. Clean water pixel selection relied on visual interpretation. The ratio of ΔR_{rc} in the red and NIR bands was used to determine whether NSB were present in the current water color anomalies. This method was used to label NSBs in CZI data. As shown in Figure 5, each step of the method relies on a large number of manual operations, making it impossible to automatically recognize NSB.

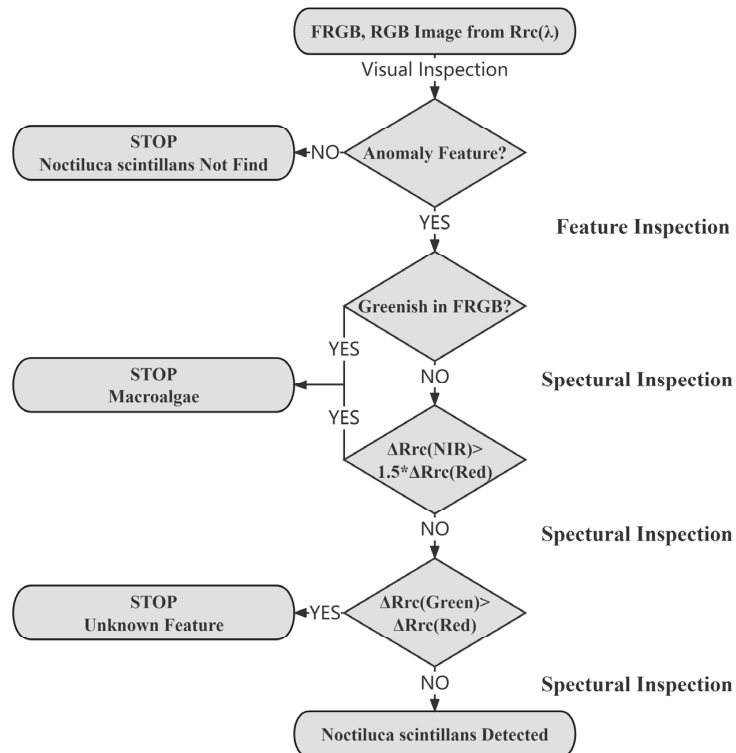


Figure 5. Flowchart of labeling the NSB region.

2.4. Dataset Construction

Considering the area ratio of NSBs and background in remote sensing images and the limited memory capacity of computing devices, the input size of NSBRNet is set to 128×128 pixels. Therefore, the annotated CZI data and the corresponding NSB labels were randomly divided into 128×128 pixel samples. After random partitioning, a total of 1108 samples of CZI data were obtained and randomly divided into a training set (60%), validation set (20%), and test set (20%) at a ratio of 6:2:2. To obtain enough training data and avoid overfitting the model, the dataset was randomly enhanced by horizontal flipping, vertical flipping, diagonal flipping, random angle rotation, and random scale scaling. Examples of data augmentation results are shown in Figure 6:

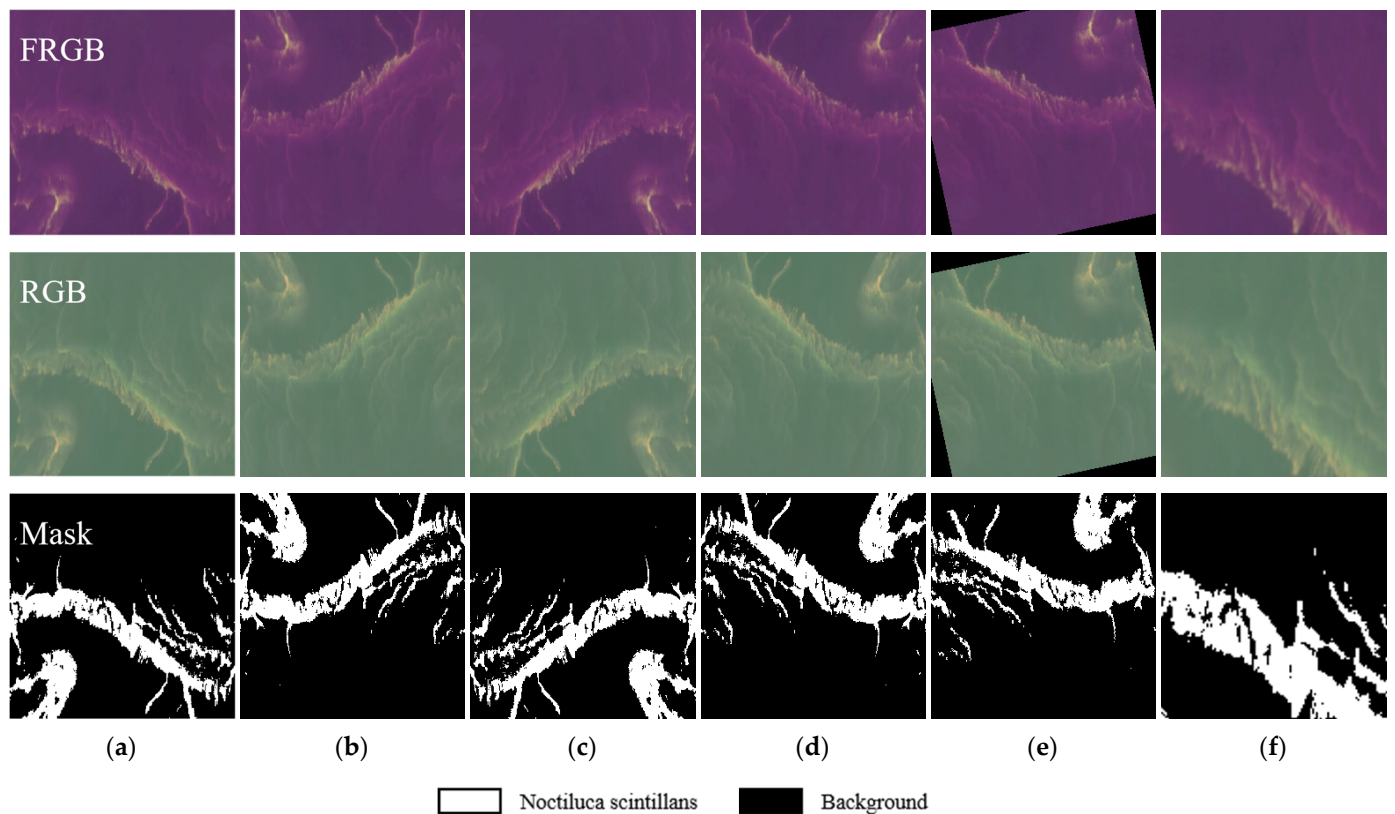


Figure 6. Examples of data augmentation results. (a) Origin; (b) horizontal flipping; (c) vertical flipping; (d) diagonal flipping; (e) random angle rotation; (f) random scale scaling.

2.5. Evaluation Criteria

The confusion matrix proposed by Congalton in 1991 [61] is the basis for algorithm validation, and almost all algorithm evaluation metrics are implemented based on the confusion matrix, which is shown in Table 2.

Table 2. Confusion matrix for binary classification.

Confusion Matrix		Ground Truth	
		Positive	Negative
Recognition	Positive	True-Positive (TP)	False-Positive (FP)
Result	Negative	False-Negative (FN)	True-Negative (TN)

Where TP and FP are the number of NSB pixels and non-NSB pixels identified as NSB, respectively; while TN and FN are the number of non-NSB pixels and NSB pixels identified as non-NSB, respectively.

In remote sensing images, the ratio of NSB area to total area is extremely small, usually less than 2%, which often leads to overall accuracy (OA) failure. This means that NSB remote sensing recognition is a serious category imbalance task. To measure the

performance of different methods, four evaluation metrics: precision, recall, F1-score, and Intersection Over Union (IOU), are used in this study, and they are calculated as follows:

$$Precision = TP / (TP + FP) \quad (5)$$

$$Recall = TP / (TP + FN) \quad (6)$$

$$F1 - Score = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (7)$$

$$IOU = TP / (TP + FP + FN) \quad (8)$$

3. *Noctiluca scintillans* Bloom Recognition Network (NSBRNet)

3.1. Network Structure

In order to achieve the automatic recognition of NSBs in remote sensing data, diverse features need to be fully explored. In this study, the Inception Conv Block and Swin Attention Block were used to develop NSBRNet. Similar to classical semantic segmentation models such as UNet, NSBRNet also adopts the same symmetric encoder-decoder structure for feature extraction and processing. The network structure of NSBRNet is shown in Figure 7. The input image is five-channel data with a size of 128×128 pixels, and the output is the NSB recognition with the same size as the input. The input data include Rrc data at four bands and NSBI to increase the difference between NSBs and background. The input will first go through 4 rounds of downsampling and hybrid feature extraction in the encoder after the preprocessing convolution layer. In the encoder, the hybrid feature extraction combined with the Inception Conv Block and Swin Attention Block encodes the different features of NSBs, while the downsampling and hybrid feature extraction stacked step by step can expand the receptive field of the model. It enables the encoder to learn the details and semantic features of NSBs at different scales from the input, which can effectively improve the generalization performance of the model to NSBs at various scales. Unlike the encoder, the hybrid feature extraction is used for feature decoding in the decoder, which decodes the final NSB location information by 4 segments of similar level-by-level feature decoding and upsampling. In addition, to reduce the loss of underlying features brought by downsampling, NSBRNet sets jump connections between the corresponding layers of the encoder-decoder structure. In the decoder, the feature maps encoded by the encoder are upsampled and concatenated with the feature maps transmitted by the jump connection, and then hybrid feature extraction is performed. This dense connection can well preserve the NSB region with detailed features. Finally, the feature information decoded by the decoder is assigned a yes/no category label of NSBs for each pixel through the postprocessing convolution layer and output as the result of NSB remote sensing recognition.

As shown in Figure 8, the hybrid feature extraction consists of two basic components: Inception Conv Block and Swin Attention Block. Convolution is proficient in extracting local high-frequency information through weight-sharing convolution kernels, while self-attention tends to dynamically calculate the similarity between different patches to obtain global low-frequency information. To fully integrate the advantages of these two different feature extraction mechanisms, inspired by the study of Si et al. [44], the channel spectral features were extracted in the hybrid feature extraction from the 1×1 convolutional layer and divided into a high-frequency part and a low-frequency part. Then, the high-frequency channel spectral features and local detail features of NSBs were extracted by the Inception Conv Block, and the low-frequency global spatial features of NSBs were extracted by the Swin Attention Block. The shallow layer of the model is usually used to capture various detailed features, while the deep layer of the model tends to target the global semantic information. As the depth of the model increases, the contribution of the Swin Attention Block to the model gradually increases; as a result, the proportion of low-frequency information increases accordingly. Considering the high spatial resolution

of CZI data (i.e., 50 m), the proportion of convolutions and self-attention in each layer of NSBRNet was set to 6:2, 5:3, 4:4, and 4:4.

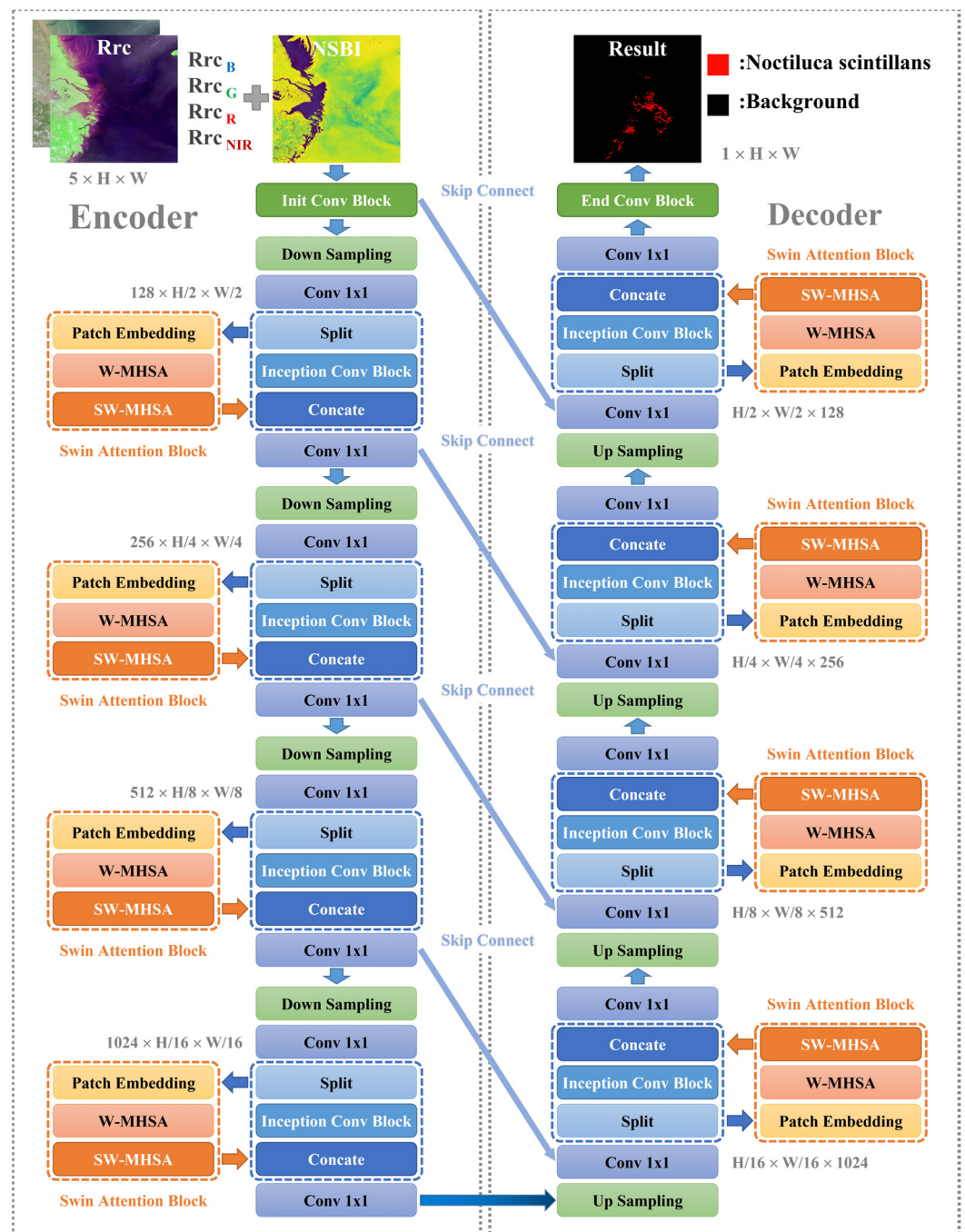


Figure 7. The model structure of NSBRNet.

3.2. Inception Conv Block

The spectral features and local detail features are essential factors in determining whether *Noctiluca scintillans* is present in the pixel, which requires the model to efficiently extract the spectral features of the NSB. At the same time, there is other prior knowledge in remote sensing images, i.e., a pixel has a great probability of being similar to its surrounding neighboring pixels.

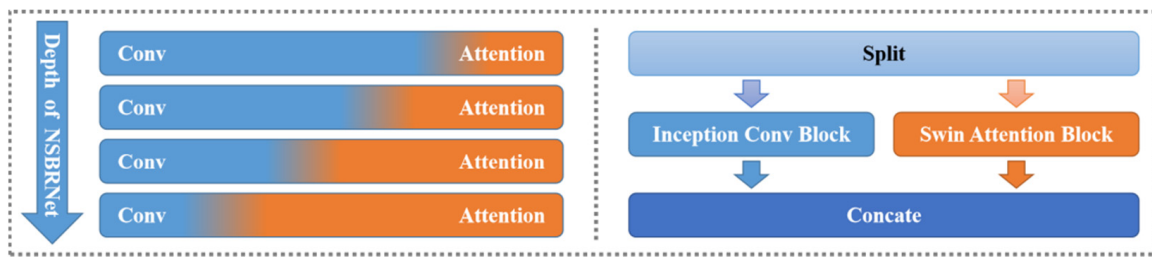


Figure 8. Structure of hybrid feature extraction.

To improve the performance of NSBRNet by using multiple features simultaneously, the Inception Conv Block (ICB) is designed in this study to capture the characteristics of NSBs in remote sensing data. As shown in Figure 9, ICB consists of three parts. The channel spectral features extracted by the 1×1 convolutional layer are first copied into three parallel branches. The first branch is the skip connect layer, which not only preserves the channel spectral features extracted by 1×1 convolution but also enables NSBRNet to deepen the network depth and improve the network nonlinear fitting ability while maintaining the ability of stable optimization of the network. This layer also slows down the degradation of the deep network and avoids overfitting and gradient disappearance of the network. The feature maps in the second branch are first subjected to 3×3 average pooling to fuse the channel spectral information of neighboring pixels, and then the fused channel spectral features are extracted using 1×1 convolution. The third branch is a conventional stacking of 3×3 convolutions to extract local features of NSBs in a small range, where the stacking of two layers of 3×3 convolutions can increase the receptive field to a 5×5 range with a smaller number of parameters than a single layer of 5×5 convolutions. Finally, the different channel features and local information in the three branches are cascaded together by the concatenation operation as the output feature map of the ICB. The ICB can be represented as:

$$\begin{aligned}
 F_{Input1}, F_{Input2}, F_{Input3} &= Duplicate(F_{Input}) \\
 F_{Local1} &= F_{Input1} \\
 F_{Local2} &= Conv_{1 \times 1}(AvgPool(F_{Input2})) \\
 F_{Local3} &= Conv_{3 \times 3}(Conv_{3 \times 3}(F_{Input3})) \\
 F_{channel} &= Concat(F_{Local1}, F_{Local2}, F_{Local3})
 \end{aligned}
 \tag{9}$$

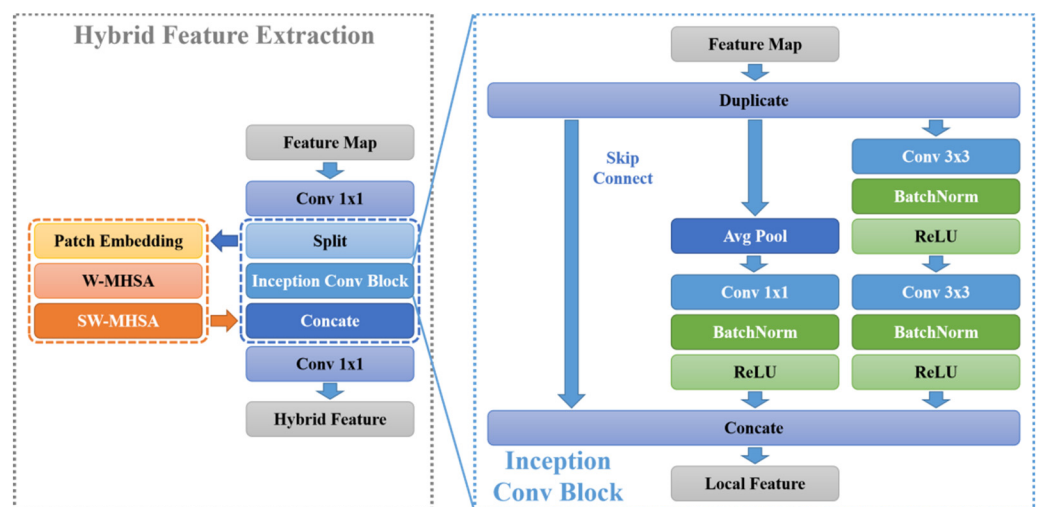


Figure 9. Structure of Inception Conv Block.

3.3. Swin Attention Block

Repeated stacking of convolutional and downsampling layers can expand the receptive field of the model, but it causes a large loss of feature information and leads to a surge in the number of model parameters. Self-attention is another peer-to-peer method in representation learning that can capture the connection between different regions in the global feature map by dynamically computing the similarity between patches. Therefore, self-attention is often used to model the correlation relationships within the data.

In the NSB remote sensing recognition task, the input feature maps are divided into non-overlapping patches by the patch embedding operation. Since the pixels with NSBs are usually clustered together, the patches containing NSB regions have strong semantic correlation among the patches. However, computing the global self-attention among all patches brings a computational complexity of quadratic number of patches, which limits the application of self-attention in recognition of NSBs. In contrast, the Swin Transformer based on the shift window strategy proposed by Liu et al. [47] restricts the self-attention computation to the inside of the window and moves the window to obtain attentional interactions across the window. As shown in Figure 10, the Swin Attention Block (SAB) consists of patch embedding, W-MHSA, and SW-MHSA. Windows multi-head self-attention (W-MHSA) and shifted windows multi-head self-attention (SW-MHSA) replace the single-layer self-attention in traditional ViT, effectively reducing the parameters and computation of the model while retaining the ability to model global long-range dependencies. Therefore, we fuse SAB with ICB in NSBRNet to provide global contextual information of NSBs across multiple patches during hybrid feature extraction, and dynamically generate codec weights of different regions in the feature map.

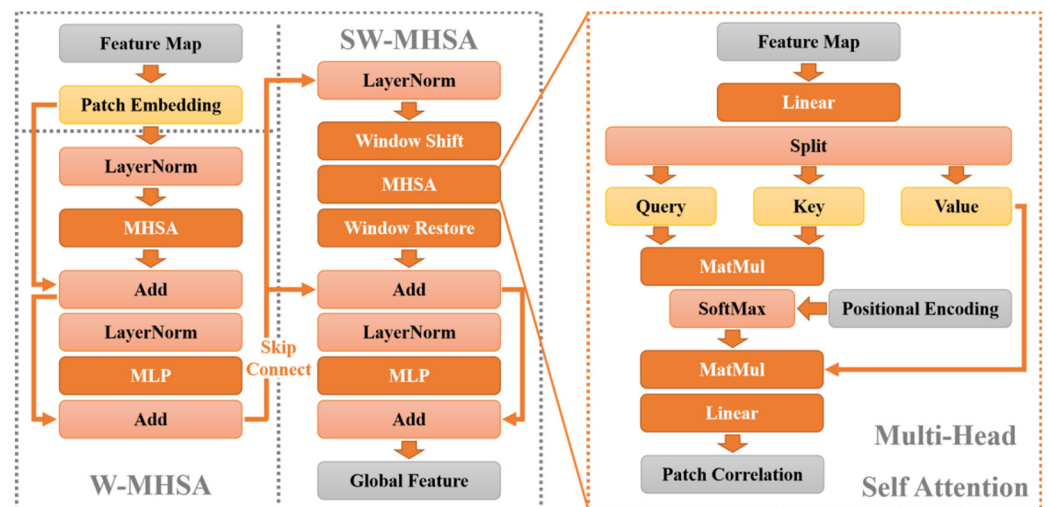


Figure 10. Structure of the Swin Attention Block.

In SAB, W-MHSA divides the input feature map into windows that do not overlap each other and uses MHSA to calculate the correlation between different patches within windows. Subsequently, in SW-MHSA, the result of self-attention within windows will be divided after displacement into new windows, at which time, the region once located at the edge of the windows becomes close to the center of the windows, and the MHSA in SW-MHSA calculates the global correlations of the feature maps using the new windows with the assistance of the location mask. SAB is useful in prompting the model to focus on the NSB regional spatial features, and it will also suppress the interference features brought by different background regions, weaken the influence of the surrounding marine environment on NSB recognition, and provide more high-level semantic information for

the model in the deeper layers of the network to solve NSB remote sensing recognition problems in complex background environments. SAB can be represented as:

$$\begin{aligned}
 F_w &= W - \text{MHSA}(\text{LN}(F_{\text{Input}})) + F_{\text{Input}} \\
 F_w &= \text{MLP}(F_w) + F_w \\
 F_{sw} &= SW - \text{MHSA}(\text{LN}(F_{sw})) + F_{sw} \\
 F_{\text{Global}} &= \text{MLP}(F_{sw}) + F_{sw}
 \end{aligned} \tag{10}$$

4. Results

4.1. Result Validation

We first examined the performances of NSBRNet using the reserved CZI data. Figure 11a shows two regions mentioned in the NSOAS report. Since only a few CZI images were available, region (A) was partially used for model training, while region (B) was not used for model training. The remote sensing recognition results of NSBs in Figure 11d demonstrate that NSBRNet shows accurate NSB recognition capability in CZI remote sensing images.

4.2. Comparison Experiment

To further measure the performance of NSBRNet, we compared it with some common deep learning models: Res UNet [62], UNet [26], Swin UNet [51], Trans UNet [50], FCN-8s [25], and PSPNet (ResNet34) [63] using the same hyperparameters and inputs. The comparison experiment was conducted on two NVIDIA RTX 3090 graphics processing units (GPUs) with 24G graphics memory, and various semantic segmentation models were implemented using the open source deep learning library PyTorch. To compare NSBRNet with other deep learning models, we controlled all experiments with the same hardware and software, and set the following identical hyperparameters for all models: batch size set to 32, training epoch set to 120, combining cross-entropy and boundary loss [64,65] as loss functions, using Adam as the optimizer, setting the initial learning rate to 1×10^{-3} , and using a cosine annealing learning rate decay strategy to make the learning rate fluctuate in a cosine-like cycle.

Table 3 lists the recognition results with various deep learning models using the same CZI data. The recall and IOU of NSBRNet are 88.20% and 82.18%, respectively, and the precision and F1-score are 92.22% and 90.10%, respectively, which achieve the best comprehensive performance in the test set. The experimental results also show that NSBRNet has strong adaptability to identify NSB in different background environments. In addition, although self-attention is good at global modeling, it lacks fine-grained localization ability. To avoid this problem, NSBRNet uses a flexible combination of convolution and self-attention by channel partitioning to extract global spatial information while preserving channel and local detail information. While Trans UNet replaces some of the convolution layers in the encoder with self-attention, Swin UNet is built entirely with self-attention, and the above self-attention addition affects the extraction of the underlying detail features by the model. In NSB recognition, the channel and detail features play a decisive role in the recognition results. Therefore, with the same hyperparameters, the results show that the recall and IOU of Res UNet and UNet are 85.42% and 79.22%, and 84.46% and 77.82%, respectively, outperforming Swin UNet and Trans UNet. The models with tandem structures such as PSP Net and FCN-8s achieved the worst results in the CZI dataset, which proves that the models with encoder-decoder structures are more suitable for identifying NSBs in remote sensing data than the deep learning models with other structures. Figure 12 shows that NSBRNet can accurately detect NSBs with CZI data. By comparing the recognition results with other models, NSBRNet clearly has obvious advantages in both NSB patches and edges. Additionally, NSBRNet can better distinguish NSBs with low biomass or in near-shore turbid water.

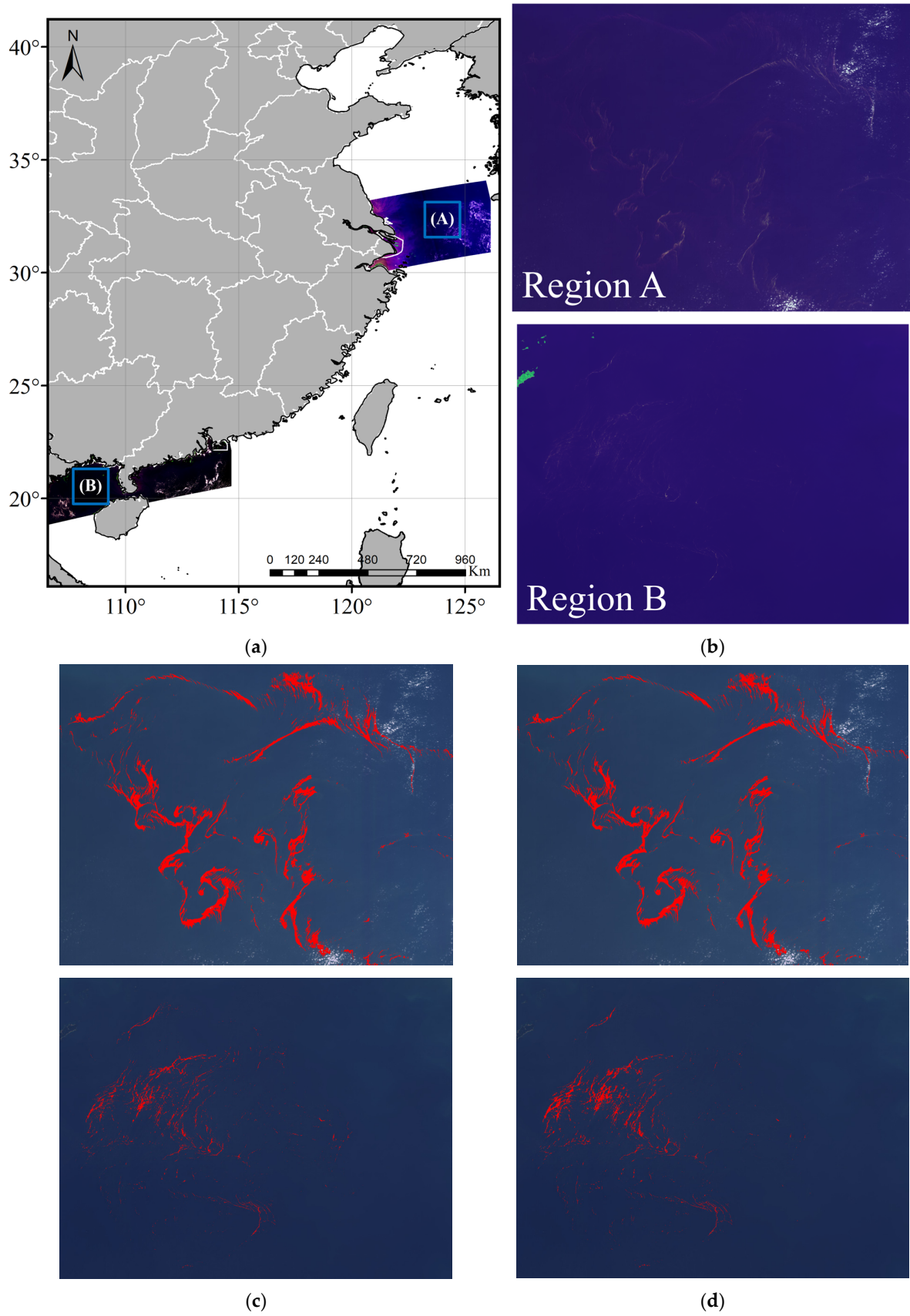
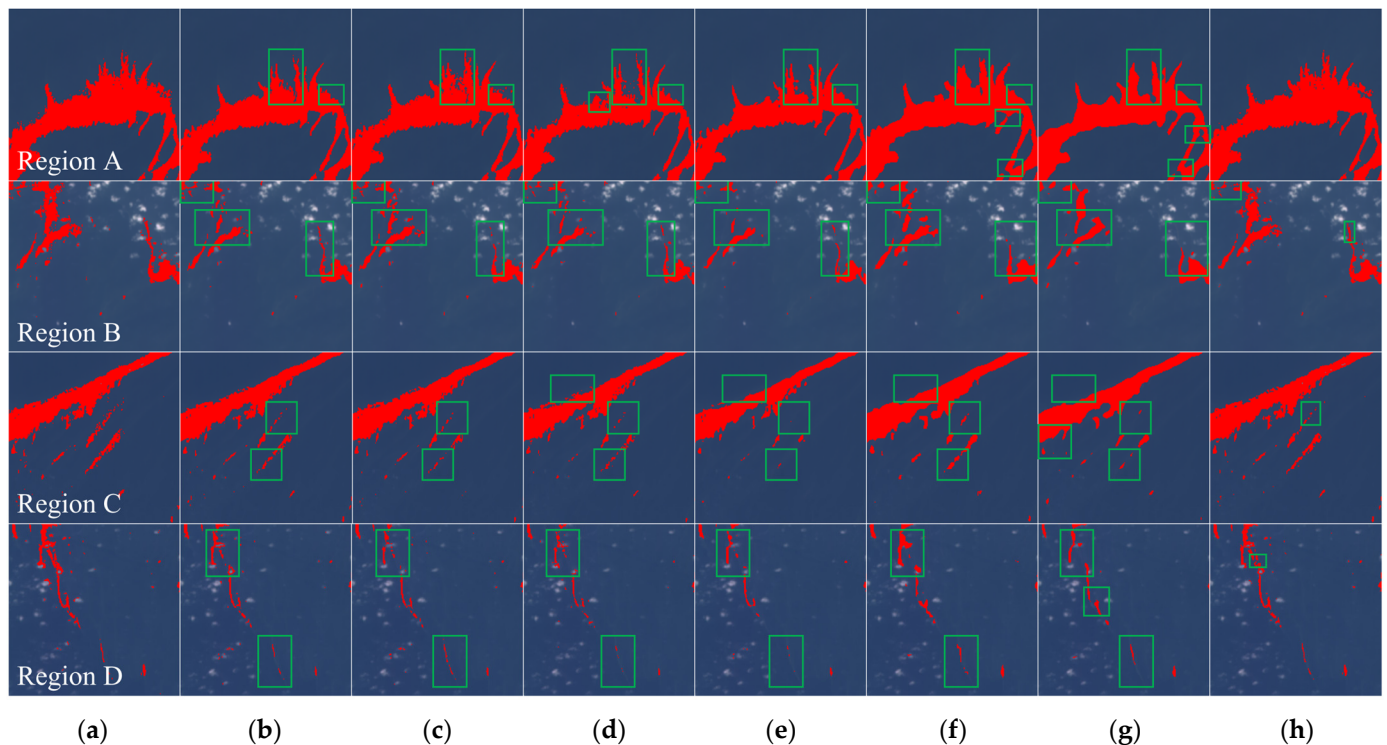


Figure 11. CZI NSB validation result. (a) Region; (b) FRGB; (c) ground truth; (d) NSBRNet.

Table 3. Performance of each model in the CZI dataset.

Model	Precision	Recall	F1-Score	IOU
Res UNet	91.51	85.42	88.26	79.22
UNet	90.75	84.46	87.38	77.82
Swin UNet	87.72	83.64	85.49	74.92
Trans UNet	87.76	82.79	85.03	74.25
FCN-8s	85.00	80.57	82.67	70.75
PSPNet (ResNet34)	84.52	76.28	80.12	66.94
NSBRNet	92.22	88.20	90.10	82.18

**Figure 12.** CZI NSB recognition result. (a) Ground truth; (b) Res UNet; (c) UNet; (d) Swin UNet; (e) Trans UNet; (f) FCN-8s; (g) PSPNet; (h) NSBRNet.

The remote sensing recognition results of NSBs for different models in the CZI data are shown in the following figures:

5. Discussion

5.1. Ablation Experiment

As the design of NSBRNet inherits the structure of UNet [26], this section conducts ablation experiments using UNet as the baseline model to evaluate the influence of ICB and SAB on the performance of NSB recognition. The same computing resources and input parameters were used in the ablation experiment.

The results of the ablation experiments are shown in Figure 13 and Table 4. Although both ICB and SAB improved the recognition performance of the NSBRNet model, ICB mainly contributes to the improvement of the model. In other words, the spectral and local detail features in the NSBs play a decisive role in the recognition results. The results show that NSBRNet incorporating ICB and SAB can achieve better performance than that of the single structure model.

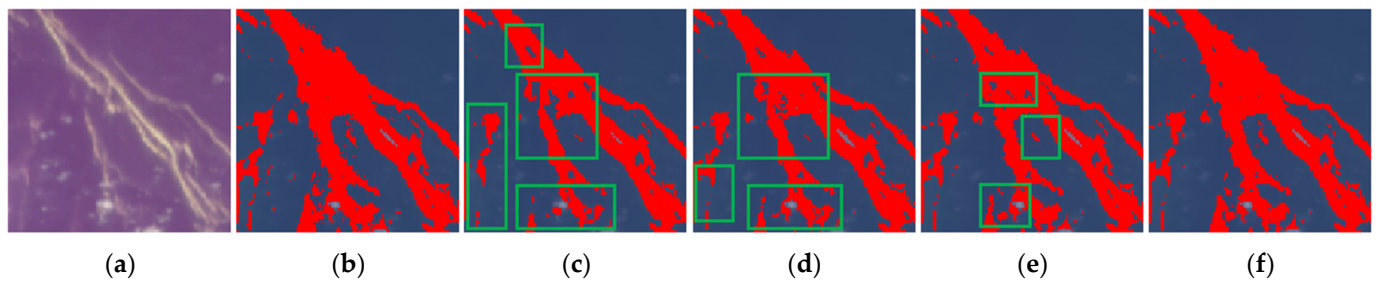


Figure 13. Influence of each model on the performance of NSB recognition. (a) FRGB; (b) ground truth; (c) UNet; (d) NSBRNet (SAB); (e) NSBRNet (ICB); (f) NSBRNet. Green frames are locations where the differences between model recognized and the ground truth are larger.

Table 4. Performance of each model in the dataset.

Model	Precision	Recall	F1-Score	IOU
UNet	90.75	84.46	87.38	77.82
NSBRNet (SAB)	91.52	85.94	88.56	79.70
NSBRNet (ICB)	92.02	86.14	88.89	80.28
NSBRNet	92.22	88.20	90.10	82.18

5.2. NSBI Applicability Analysis

To evaluate the effect of NSBI on the performance of NSBRNet, we compared the results with 4 channels (R, G, B, NIR) and 5 channels (R, G, B, NIR, NSBI) as the input of NSBRNet. Similar to the ablation experiment, the same computing resources and input parameters were used in the NSBI applicability analysis.

Adding one more channel in the input, the accuracy of the NSBRNet model slightly improved as shown in Table 5 and Figure 14. For example, recall slightly increased from 87.25% to 88.20% and IOU increased from 81.25% to 82.18% when NSBI data were included in the model. The results show that NSBI can still improve the NSB recognition performance of NSBRNet to some extent.

Table 5. Impact of NSBI on the performance of NSBRNet.

Model	Precision	Recall	F1-Score	IOU
R, G, B, NIR	92.11	87.25	89.51	81.25
R, G, B, NIR, NSBI	92.22	88.20	90.10	82.18

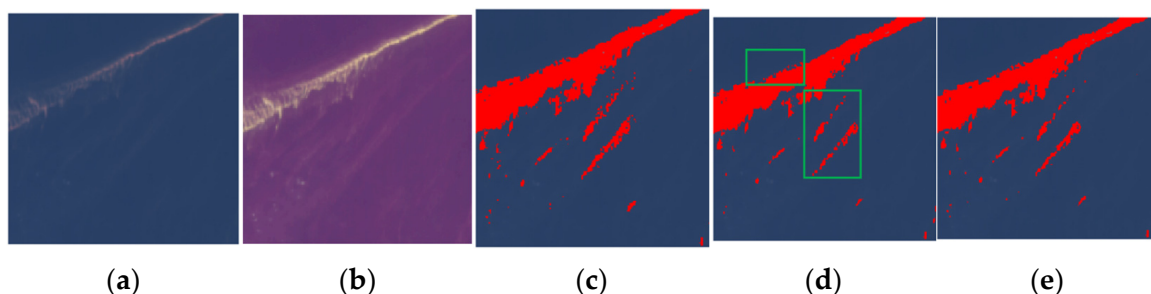


Figure 14. Influence of NSBI on the performance of NSB recognition. (a) RGB; (b) FRGB; (c) ground truth; (d) R, G, B, NIR; (e) R, G, B, NIR, NSBI.

6. Conclusions

In this study, we developed a remote sensing recognition model NSBRNet incorporating an Inception Conv Block and a Swin Attention Block based on the latest deep learning technology using CZI data, where ICB uses convolution to extract channel and local detail

features, and SAB uses self-attention to extract global spatial features. NSBRNet significantly combines the advantages of convolution and self-attention through the channel partitioning mechanism, which extracts the global spatial information while preserving the channel and local detail information, making NSBRNet flexible in constructing and fitting most of the information compared with other models. The hybrid feature of NSBRNet enables the model to delineate NSB features at different spatial scales and different background conditions. To improve the recognition accuracy, we developed NSBI and used it as an input to the model.

We collected the NSB event reports released by NSOAS to check the performance of the model. Validation results showed that the model can efficiently and automatically recognize NSB from CZI images. Comparison and ablation experiments were then conducted to compare the performance of the developed NSB model with other deep learning models and to examine the performance of ICB, SAB, and NSBI in the NSB model. In the comparison experiments, NSBRNet performed better with an IOU of 82.18% and recall of 88.20%, which were improved by 2.96–15.24% and 2.78–11.92% compared to other deep learning models. In the ablation experiment, ICB and SAB had 2.46% and 1.88% improvements in IOU compared to baseline UNet, respectively, and NSBI had a 0.93% improvement in IOU for NSBRNet.

The NSBRNet proposed in this study can be quickly extended to many other satellite sensors with higher spatial resolution such as Landsat OLI (~30 m) and Sentinel MSI (~10 m), or lower spatial resolution but with wide swath such as MODIS (~250 m) and OLCI (~300 m) by constructing the corresponding datasets. By combining different satellite data, the global appearance of RNS can be well observed from space and their spatio-temporal characteristics can be comprehensively analyzed. This will provide useful information in the future research to study the relationships between RNS bloom and the relevant environmental factors.

Author Contributions: Conceptualization, S.C.; methodology, H.C. (Hanlin Cui); validation, H.C. (Hanlin Cui); data curation, H.C. (Haobin Cai); writing—original draft preparation, H.C. (Hanlin Cui); writing—review and editing, L.H., S.C., J.W., H.C. (Hanlin Cui), C.M., J.L. and B.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (NSFC), grant number T2222010.

Data Availability Statement: Not applicable.

Acknowledgments: Thanks to the National Satellite Ocean Application Service (NSOAS) for providing the HY1C/D remote sensing satellite data for this research. We would like to thank three anonymous reviewers for their helpful comments and constructive suggestions, which have significantly improved the manuscript.

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

References

1. Elbrächter, M.; Qi, Z. Aspects of *Noctiluca* (Dinophyceae) population dynamics. In *Physiological Ecology of Harmful Algal Blooms*; Anderson, D.M., Cembella, A.D., Hallegraeff, G.M., Eds.; Springer: Berlin/Heidelberg, Germany, 1998; pp. 315–335, ISBN 3540641173.
2. Tang, D.; Di, B.; Wei, G.; Ni, I.-H.; Oh, I.S.; Wang, S. Spatial, seasonal and species variations of harmful algal blooms in the South Yellow Sea and East China Sea. *Hydrobiologia* **2006**, *568*, 245–253. [[CrossRef](#)]
3. Harrison, P.J.; Furuya, K.; Glibert, P.M.; Xu, J.; Liu, H.B.; Yin, K.; Lee, J.H.W.; Anderson, D.M.; Gowen, R.; Al-Azri, A.R.; et al. Geographical distribution of red and green *Noctiluca scintillans*. *Chin. J. Oceanol. Limnol.* **2011**, *29*, 807–831. [[CrossRef](#)]
4. Song, J.; Bi, H.; Cai, Z.; Cheng, X.; He, Y.; Benfield, M.C.; Fan, C. Early warning of *Noctiluca scintillans* blooms using in-situ plankton imaging system: An example from Dapeng Bay, PR China. *Ecol. Indic.* **2020**, *112*, 106123. [[CrossRef](#)]
5. Huang, C.; Qi, Y. The abundance cycle and influence factors on red tide phenomena of *Noctiluca scintillans* (Dinophyceae) in Dapeng Bay, the South China Sea. *J. Plankton Res.* **1997**, *19*, 303–318. [[CrossRef](#)]
6. Do Rosário Gomes, H.; Goes, J.I.; Matondkar, S.P.; Buskey, E.J.; Basu, S.; Parab, S.; Thoppil, P. Massive outbreaks of *Noctiluca scintillans* blooms in the Arabian Sea due to spread of hypoxia. *Nat. Commun.* **2014**, *5*, 4862. [[CrossRef](#)] [[PubMed](#)]

7. Buskey, E.J. Growth and bioluminescence of *Noctiluca scintillans* on varying algal diets. *J. Plankton Res.* **1995**, *17*, 29–40. [[CrossRef](#)]
8. Xue, C.; Chen, S.; Zhang, T. Optical proxy for the abundance of red *Noctiluca scintillans* from bioluminescence flash kinetics in the Yellow Sea and Bohai Sea. *Opt. Express* **2020**, *28*, 25618–25632. [[CrossRef](#)] [[PubMed](#)]
9. Rohr, J.; Hyman, M.; Fallon, S.; Latz, M.I. Bioluminescence flow visualization in the ocean: An initial strategy based on laboratory experiments. *Deep Sea Res. Part I* **2002**, *49*, 2009–2033. [[CrossRef](#)]
10. Lapota, D. Night time surveillance of harbors and coastal areas using bioluminescence camera and buoy systems. In Proceedings of the Photonics for Port and Harbor Security, Orlando, FL, USA, 29–30 March 2005; pp. 128–137.
11. Schaumann, K.; Gerdes, D.; Hesse, K. Hydrographic and biological characteristics of a *Noctiluca scintillans* red tide in the German Bight, 1984. *Meeresforschung* **1988**, *32*, 77–91. [[CrossRef](#)]
12. Uhlig, G.; Sahling, G. Long-term studies on *Noctiluca scintillans* in the German Bight population dynamics and red tide phenomena 1968–1988. *Neth. J. Sea Res.* **1990**, *25*, 101–112. [[CrossRef](#)]
13. Tseng, L.-C.; Kumar, R.; Chen, Q.-C.; Hwang, J.-S. Summer distribution of *Noctiluca scintillans* and mesozooplankton in the Western and Southern East China Sea prior to the Three Gorges Dam operation. *Hydrobiologia* **2011**, *666*, 239–256. [[CrossRef](#)]
14. Junwu, T.; Jing, D.; Qimao, W.; Chaofei, M. Research of the effects of atmospheric scattering on red tide remote sensing with normalized vegetation index. *Acta Oceanol. Sin.* **2004**, *26*, 136–142. (In Chinese)
15. Hu, C.; Muller-Karger, F.E.; Taylor, C.J.; Carder, K.L.; Kelble, C.; Johns, E.; Heil, C.A. Red tide detection and tracing using MODIS fluorescence data: A regional example in SW Florida coastal waters. *Remote Sens. Environ.* **2005**, *97*, 311–321. [[CrossRef](#)]
16. Ahn, Y.-H.; Shanmugam, P. Detecting the red tide algal blooms from satellite ocean color observations in optically complex Northeast-Asia Coastal waters. *Remote Sens. Environ.* **2006**, *103*, 419–437. [[CrossRef](#)]
17. Takahashi, W.; Kawamura, H.; Omura, T.; Furuya, K. Detecting red tides in the eastern Seto inland sea with satellite ocean color imagery. *J. Oceanogr.* **2009**, *65*, 647–656. [[CrossRef](#)]
18. Sakuno, Y.; Maeda, A.; Mori, A.; Ono, S.; Ito, A. A simple red tide monitoring method using sentinel-2 data for sustainable management of Brackish Lake Koyama-ike, Japan. *Water* **2019**, *11*, 1044. [[CrossRef](#)]
19. Qi, L.; Tsai, S.F.; Chen, Y.; Le, C.; Hu, C. In Search of Red *Noctiluca scintillans* Blooms in the East China Sea. *Geophys. Res. Lett.* **2019**, *46*, 5997–6004. [[CrossRef](#)]
20. Qi, L.; Hu, C.; Liu, J.; Ma, R.; Zhang, Y.; Zhang, S. *Noctiluca* blooms in the East China Sea bounded by ocean fronts. *Harmful Algae* **2022**, *112*, 102172. [[CrossRef](#)] [[PubMed](#)]
21. Dwivedi, R.; Priyaja, P.; Rafeeq, M.; Sudhakar, M. MODIS-Aqua detects *Noctiluca scintillans* and hotspots in the central Arabian Sea. *Environ. Monit. Assess.* **2016**, *188*, 1–11. [[CrossRef](#)]
22. Liu, R.-J.; Zhang, J.; Cui, B.-G.; Ma, Y.; Song, P.-J.; An, J.-B. Red tide detection based on high spatial resolution broad band satellite data: A case study of GF-1. *J. Coast. Res.* **2019**, *90*, 120–128. [[CrossRef](#)]
23. Liu, R.; Xiao, Y.; Ma, Y.; Cui, T.; An, J. Red tide detection based on high spatial resolution broad band optical satellite data. *ISPRS J. Photogramm. Remote Sens.* **2022**, *184*, 131–147. [[CrossRef](#)]
24. Zhao, X.; Liu, R.; Ma, Y.; Xiao, Y.; Ding, J.; Liu, J.; Wang, Q. Red Tide Detection Method for HY-1D Coastal Zone Imager Based on U-Net Convolutional Neural Network. *Remote Sens.* **2021**, *14*, 88. [[CrossRef](#)]
25. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
26. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
27. 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](#)] [[PubMed](#)]
28. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
29. Lee, M.-S.; Park, K.-A.; Chae, J.; Park, J.-E.; Lee, J.-S.; Lee, J.-H. Red tide detection using deep learning and high-spatial resolution optical satellite imagery. *Int. J. Remote Sens.* **2019**, *41*, 5838–5860. [[CrossRef](#)]
30. Kim, S.M.; Shin, J.; Baek, S.; Ryu, J.-H. U-Net convolutional neural network model for deep red tide learning using GOCI. *J. Coast. Res.* **2019**, *90*, 302–309. [[CrossRef](#)]
31. 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, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
32. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
33. 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, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
34. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.

35. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
36. Liu, J.; Li, C.; Liang, F.; Lin, C.; Sun, M.; Yan, J.; Ouyang, W.; Xu, D. Inception convolution with efficient dilation search. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 11486–11495.
37. Lin, Z.; Feng, M.; dos Santos, C.N.; Yu, M.; Xiang, B.; Zhou, B.; Bengio, Y. A structured self-attentive sentence embedding. *arXiv* **2017**, arXiv:1703.03130.
38. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S. An image is worth 16×16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
39. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3146–3154.
40. Huang, Z.; Wang, X.; Huang, L.; Huang, C.; Wei, Y.; Liu, W. Ccnet: Criss-cross attention for semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 603–612.
41. Wu, H.; Xiao, B.; Codella, N.; Liu, M.; Dai, X.; Yuan, L.; Zhang, L. Cvt: Introducing convolutions to vision transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 22–31.
42. Dai, Z.; Liu, H.; Le, Q.V.; Tan, M. Coatnet: Marrying convolution and attention for all data sizes. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 3965–3977. [[CrossRef](#)]
43. Li, K.; Wang, Y.; Gao, P.; Song, G.; Liu, Y.; Li, H.; Qiao, Y. Uniformer: Unified transformer for efficient spatiotemporal representation learning. *arXiv* **2022**, arXiv:2201.04676.
44. Si, C.; Yu, W.; Zhou, P.; Zhou, Y.; Wang, X.; Yan, S. Inception Transformer. *arXiv* **2022**, arXiv:2205.12956.
45. Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Torr, P.H. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 6881–6890.
46. Wang, W.; Xie, E.; Li, X.; Fan, D.-P.; Song, K.; Liang, D.; Lu, T.; Luo, P.; Shao, L. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 568–578.
47. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
48. Gao, Y.; Zhou, M.; Metaxas, D.N. UTNet: A hybrid transformer architecture for medical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Online, 27 September–1 October 2021; pp. 61–71.
49. Sha, Y.; Zhang, Y.; Ji, X.; Hu, L. Transformer-Unet: Raw Image Processing with Unet. *arXiv* **2021**, arXiv:2109.08417.
50. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv* **2021**, arXiv:2102.04306.
51. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-unet: Unet-like pure transformer for medical image segmentation. *arXiv* **2021**, arXiv:2105.05537.
52. Yuan, J.; Wang, L.; Cheng, S. STransUNet: A Siamese TransUNet-Based Remote Sensing Image Change Detection Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 9241–9253. [[CrossRef](#)]
53. Li, Q.; Zhong, R.; Du, X.; Du, Y. TransUNetCD: A Hybrid Transformer Network for Change Detection in Optical Remote-Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 3169479. [[CrossRef](#)]
54. Zhang, C.; Wang, L.; Cheng, S.; Li, Y. SwinSUNet: Pure Transformer Network for Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 3160007. [[CrossRef](#)]
55. Zhang, C.; Jiang, W.; Zhang, Y.; Wang, W.; Zhao, Q.; Wang, C. Transformer and CNN Hybrid Deep Neural Network for Semantic Segmentation of Very-High-Resolution Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 3144894. [[CrossRef](#)]
56. He, X.; Zhou, Y.; Zhao, J.; Zhang, D.; Yao, R.; Xue, Y. Swin Transformer Embedding UNet for Remote Sensing Image Semantic Segmentation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 3144165. [[CrossRef](#)]
57. Yao, J.; Jin, S. Multi-Category Segmentation of Sentinel-2 Images Based on the Swin UNet Method. *Remote Sens.* **2022**, *14*, 3382. [[CrossRef](#)]
58. Wang, L.; Li, R.; Zhang, C.; Fang, S.; Duan, C.; Meng, X.; Atkinson, P.M. UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS J. Photogramm. Remote Sens.* **2022**, *190*, 196–214. [[CrossRef](#)]
59. Hu, C.; Chen, Z.; Clayton, T.D.; Swarzenski, P.; Brock, J.C.; Muller-Karger, F.E. Assessment of estuarine water-quality indicators using MODIS medium-resolution bands: Initial results from Tampa Bay, FL. *Remote Sens. Environ.* **2004**, *93*, 423–441. [[CrossRef](#)]
60. Hu, C. A novel ocean color index to detect floating algae in the global oceans. *Remote Sens. Environ.* **2009**, *113*, 2118–2129. [[CrossRef](#)]
61. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46. [[CrossRef](#)]

62. Xiao, X.; Lian, S.; Luo, Z.; Li, S. Weighted res-unet for high-quality retina vessel segmentation. In Proceedings of the 2018 9th International Conference on Information Technology in Medicine and Education (ITME), Hangzhou, China, 19–21 October 2018; pp. 327–331.
63. 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, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
64. Bokhovkin, A.; Burnaev, E. Boundary loss for remote sensing imagery semantic segmentation. In Proceedings of the International Symposium on Neural Networks, Moscow, Russia, 10–12 July 2019; pp. 388–401.
65. Kervadec, H.; Bouchtiba, J.; Desrosiers, C.; Granger, E.; Dolz, J.; Ayed, I.B. Boundary loss for highly unbalanced segmentation. In Proceedings of the International Conference on Medical Imaging with Deep Learning, London, UK, 8–10 July 2019; pp. 285–296.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.