IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 11, 2022, accepted May 30, 2022, date of publication June 7, 2022, date of current version June 13, 2022.

*Digital Object Identifier 10.1109/ACCESS.2022.3180331*

# MODIS Green-Tide Detection With a Squeeze and Excitation Oriented Generative Adversarial Network

**XIFANG JIN[1,2,3], YUN LI[1], JIANHUA WAN[1], XINRONG LYU[1], PENG REN[1], (Senior Member, IEEE), AND JIE SHANG[2,3]**

[1]College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China
[2]Shandong Provincial Key Laboratory of Marine Ecological Environment and Disaster Prevention and Mitigation, Qingdao 266061, China
[3]North China Sea Marine Forecasting Center of State Oceanic Administration, Qingdao 266061, China

Corresponding author: Jianhua Wan (19850014@upc.edu.cn)

This work was supported in part by the National Key Research and Development Plan under Grant 2017YFC1405303, 2016YFC1402103.

**ABSTRACT** This paper presents a novel framework combining spectral analysis and machine learning for green-tide detection. The framework incorporates a squeeze and excitation (SE) attention module into a U-shaped generator of a generative adversarial network (GAN), and is referred to as squeeze and excitation oriented generative adversarial network (SE-GAN). In the SE-GAN, the normalized differential vegetation index (NDVI) is used as the preprocessing filter, which enhances the information associated with green-tide. The SE attention module recalibrates the feature maps so as to enhance the useful features conveyed from the generator's convolution layer while suppressing less useful ones. Overall, the generator attempts to render images that contain green-tide in a way that highly approximates the reference images, while the discriminator characterizes the difference between the generated images and the reference images. The training scheme, which is adversarial, minimizes the *f*-divergence between the generator and discriminator. Consequently, compared to other green-tide detection algorithms only applicable in small-area scenes, SE-GAN can automatically detect green-tide in MODIS images of any size. Experiments with both large- and small-format MODIS imagery confirm that SE-GAN's detection results are superior to those of five other commonly used methods.

**INDEX TERMS** GAN, green-tide detection, NDVI, SE, SE-GAN, sea-land separation.

## I. INTRODUCTION

Green-tide is a type of ecological disaster caused by the explosive proliferation of large macroalgae under specific environmental conditions. Since the 1960s and 1970s, large-scale and frequent outbreaks of green-tides have occurred in the coastal waters of many countries [1], including the largest green-tide in Yellow Sea [2], [3]. Green-tide is non-toxic, but it shields sunlight, consumes enormous amounts of oxygen and nutrients from the water, and seriously affects the growth of other marine creatures. Dead and rotting green-tide algae accumulate on the water surface causing water deterioration and producing harmful gases. Alive or dead, the chemical substances secreted by the algae have adverse effects on marine life. Outbreaks severely influence tourism and water sports, and removing green-tides requires a tremendous amount of manpower and financial resources, creating a huge burden for governments. In short, large-scale outbreaks of green-tides are responsible for immense ecological, environmental, and economic losses [3], [4].

Schreyers *et al.* [5] illustrate that the outbreaks of green-tide are caused by many factors including climate change and water eutrophication. The most massive outbreak on record of the *Ulva prolifera* green-tides in 2021 has attracted great attention [6]. Compared with traditional monitoring methods, satellite remote sensing can obtain the real-time and large-scale monitoring data with less labor and time cost, and is currently the most widely used and effective method in green-tide monitoring. How to interpret

The associate editor coordinating the review of this manuscript and approving it for publication was Stefania Bonafoni.

green-tide in satellite images more accurately becomes the focus of attention, and researchers have carried out extensive and in-depth researches in terms of the monitoring images and monitoring methods.

MODIS (Moderate-resolution Imaging Spectroradiometer) data has the characteristics of wide band range, fast data acquisition frequency, wide monitoring range, free and easy access. They are the most commonly used data in green-tide operational monitoring, and a large amount of data has been accumulated over the years.

The MODIS green-tide extraction algorithms are generally based on the spectral characteristics of green-tide which have strong absorption characteristics in the visible red band and reflection peaks in the near-infrared band. At the beginning, Gu *et al.* [7] proposed the threshold method detecting green-tides by manually setting a fixed threshold higher than the seawater. Prasetyo *et al.* [8] developed the threshold algorithm by combining the threshold method with methods for detecting color features in HSV space and made the detection more accurate. Meanwhile, several vegetation indices, such as NDVI, EVI, ARVI, RVI and DVI, were proposed to detect the *Ulva prolifera* green-tide. Of all five vegetation indices, NDVI was proved to be the most sensitive in all development phases of the green-tide in Yellow Sea and East China Sea [9]–[11].

However, the NDVI method is highly susceptible to the imaging influences of different atmospheric conditions, observation angle, water quality environment and marine environment, leading to uncertainties in the monitoring results. It is also found that deviation exist in the MODIS green-tide monitoring results using NDVI method [12], [13].

In order to improve the MODIS monitoring efficiency and accuracy of green-tide, researchers have made three attempts. The first is to propose a multi-source data fusion monitoring solution, which has higher data requirements. Ma *et al.* [13] combined optical and microwave data to analyze the time and space of the green-tide in the Yellow Sea in 2021. Some researchers proposed improved algorithms with higher accuracy based the NDVI algorithm. Hu *et al.* [14] proposed the floating algae index method (FAI), which is more accurate and stable than the NDVI method. However, it needs the short-wave infrared band and is more suitable for high-resolution images. Shi *et al.* [15] proposed the normalized algae index method (NDAI) which implements atmospheric correction based on MODIS data, but it has little effect on green-tides extraction in complex marine environment scenarios. Other researchers explored machine-learning solutions. Dong *et al.* [16] developed a pixel-by-pixel manner of predicting green-tide based on support vector machines, while Wu *et al.* [17] used a semi-supervised clustering method based on a limited amount of labeled data and mostly unlabeled data. Dogliotti *et al.* [18] combined a floating algal index, the conditions of the red band, and the color space of the image to detect floating plants close to green-tide. Notably, Qu and Dong [19] constructed a dataset containing images of green-tide with a clustering method. In other machine learning approaches, Xie *et al.*'s [20] solution involves object-oriented random forest classification and multi-scale segmentation, whereas Wang *et al.* [21] exploited deep learning with a convolution neural network that classifies patches of the image obtained by super pixels. Yin *et al.* [22] used a fully convolutional network that extracts high-level features and predicts features at the pixel level.

The aforementioned detection methods such as data fusion and the threshold method based vegetation indices either have high data requirements or have low detection efficiency in facing with the complex scenes of large-scale images. Traditional machine learning methods mentioned before have poor accuracy or low efficiency, and the CNN method requires a large amount of data. Besides, interferences such as cloud can bring obstacles to green-tide extraction. In green-tide images, the existence of green-tide has discontinuous, scattered, tiny and complex distribution characteristics, which determine that the green-tide detection should be based on large-format images and pay attention to the local distribution characteristics at the same time in order to grasp the overall distribution of the scattered green-tide more accurately. Motivated by this, we proposed an ensemble approach based on spectral analysis and machine learning. We aim at addressing the inefficiencies, inaccuracies, and large data volumes of previous approaches and achieving accurate green-tide detection of large-format images, a SE-GAN framework is proposed. The SE attention mechanism is embedded in the GAN to selectively emphasizes the useful features conveyed in the generated images while suppressing the less useful ones. The SE module is configured within a U-shaped structure generator to help improve channel performance, and the model is built via repeated adversarial training of the generator and the discriminator. The training objective of GAN is to minimize $f$-divergence so as to achieve the minimal difference between the generated detection image and the reference image. Once training is completed, the model is able to automatically detect green-tide in a MODIS image.

Given the above, the contributions of this paper include:

(a) An approach enhancing MODIS images of green-tide by applying NDVI and sea-land separation [23], [24] is framed.

(b) An SE attention module that is embedded in a GAN to minimize $f$-divergence between the generated and reference images of green-tide is presented.

(c) A novel framework, i.e., SE-GAN, for automatically detecting green-tide in any MODIS image is developed.

## II. DATA AND MATERIALS

### A. STUDY AREA

We chose the Yellow Sea as our study area. Green tide in Yellow Sea is the largest ecological disaster, which has the characteristics of large distribution and cross sea area. And it is also the most complex of all green tide disasters, which originated in the East China Sea and drifted to the southern part of Shandong Province in the North China Sea. In terms

of monitoring areas, compared with the North China Sea, the water in the East China Sea is more turbid, so detection on different underlying surfaces is needed in large-scale monitoring sea. Meanwhile, high-precision green tide detection is required for the salvage and disposal of green tide in local waters. In order to better show our experimental results, under the premise of taking into account the large-scale, cross-sea areas and high-precision for local areas, we set the experimental scope in the zone of (120°-123°E, 34°-37°N) (Fig. 1).

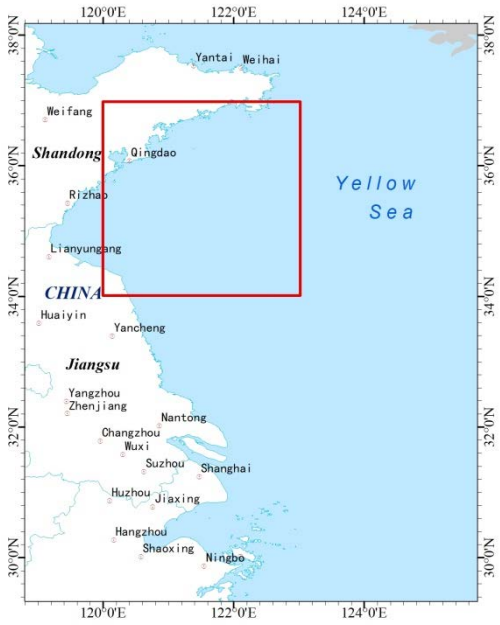



**FIGURE 1.** The study area (red rectangle showing the experiment area).

### B. SATELLITE DATA

MODIS satellite data was used in this study. It was most widely used in operational monitoring of green tide in Yellow Sea.

In this paper, we selected about 90 images of different growth stages of the green tide of 2019 for preliminary experiments, and finally chose four types of images under different scenes. First scene showed the large-scale monitoring of green tide with more green tide and fewer clouds, three other scenes showed different complicated conditions with more clouds and less green tide, including thick clouds, thick and thin clouds, thin clouds in local sea areas.

## III. METHODOLOGY

### A. TWO-STEP IMAGE FILTER PREPROCESSING

ENVI software (Esri China Information Technology Co., Ltd, Beijing, China) was used to perform basic image preprocessing and two-step image filter. Basic image preprocessing included geometric correction, atmospheric correction, radiometric calibration, etc. Two-step easy image filter before detection was designed to improve result accuracy and efficiency without increasing the operational complexity.

The two-step preprocessing filters are NVDI calculating and sea-land separation. First, green-tide has the spectral characteristics of vegetation. Compared with seawater, its spectrum produces absorption valleys in the red band and high reflection peaks in the near-infrared band [25]. This gives rise to an often-used method of monitoring for green-tide on the water's surface called the multi-band ratio method. This method enlarges the difference between the absorption valleys and the reflection peaks to highlight the information associated with green-tide. The method also helps to eliminate noise. The algorithm that extracts vegetation based on the multi-band ratio is known as the vegetation index (VI) and its variant, the normalized difference vegetation index (NDVI), has the added benefit of better reducing the influence of the atmosphere and clouds. In short, NDVI is a common and widely used vegetation index, that identifies vegetation information through the absorption and reflection of chlorophyll. Although the floating algae index (FAI) may be able to achieve better results, the NDVI method is simpler than the FAI method and the effect is not inferior to the FAI method. Xing *et al.* [26] used NDVI as an indicator and set different NDVI values to monitor green-tide in the Yellow and East China Seas. El-Gammal *et al.* [27] employed NDVI to detect vegetation cover. And Larson and Tuor [28] used NDVI to determine the time period over which vegetation was active.

NDVI is calculated with the following formula:

$$NDVI = (NIR - R)/(NIR + R). \tag{1}$$

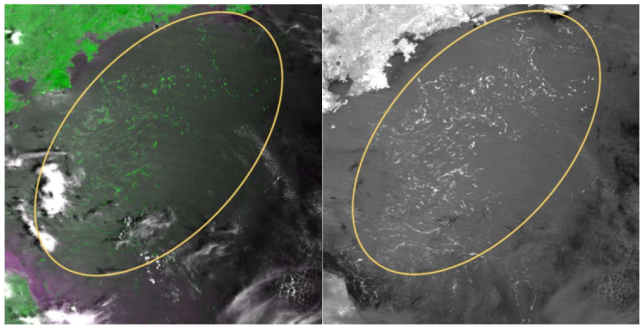
where $NIR$ is the reflectance of the near infrared band and $R$ is the reflectance of the visible red band.

The MODIS green-tide images used in this paper are multi-band images. Band 2 is the near infrared band and Band 1 is the visible red band. Hence, Eq.1 can be used to calculate the NDVI of each image. Take Fig. 2(a) as an example. Here, we used Bands 1 and 2 to synthesize the color of the green-tide so as to indicate the areas of outbreak. The image was then converted to grayscale, as shown in Fig. 2(b), and the grayscale image was enhanced to be clearer than the original green-tide image. As shown, the white green-tide on the sea in the grayscale image is more obvious than the green in Fig. 2(a) and, further, the clouds have also been eliminated.

In the second pre-processing step, we address the difficulties with distinguishing between the gray values of the green-tide and the land. This is done using the global-land-mask package from [29] and matching the latitude and longitude value of each point to determine whether that location is land or sea. Fig. 3(a) shows the land mask and Fig. 3(b) shows the result of the masking.

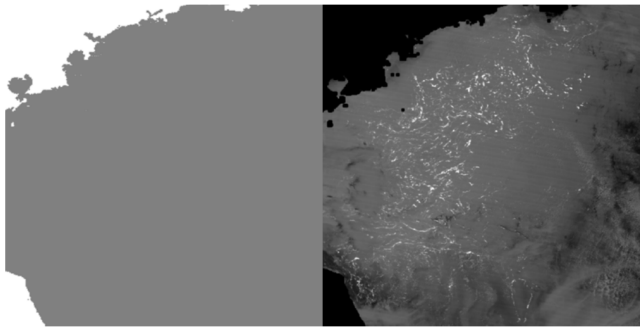
### B. THE SE-GAN FRAMEWORK

GANs [23] have been widely used in many fields. However, in recent years, GANs have seen the addition of attention to their standard architecture of a generator and a discriminator. This has significantly improved the performance of many algorithms. In this vein, SE-GAN uses squeeze and excitation



(a) Green-tide image (b) Grayscale image

**FIGURE 2.** NDVI image enhancement. The left image is the original green tide image synthesized to show color through the near infrared band (Band 2) and the visible red band (Band 1). The green tide in the original image is not as obvious as in the grayscale image on the right due to clouds and other interference.



(a) Mask Image (b) Separation Image

**FIGURE 3.** Separating land from sea. The left image is the mask derived from the global-land-mask package. It is based on the latitudes and longitudes of known land masses. The right image is the separation image after the mask has been applied.

attention to substantially improve green tide detection. This section details the SE-GAN framework.

### 1) THE SE-GAN ARCHITECTURE

SE-GAN consists of two important components, a generator and a discriminator. The generator is designed to generate a detection image that is as accurate as the reference image, while the discriminator is used to characterize the difference between the generated detection image and the reference image. Through adversarial training with the objective of minimizing $f$-divergence, the framework minimizes the difference between the generated detection image and the reference image.

Both the generator and the discriminator are composed of encoders and decoders [31]. In the generator, a convolutional layer, a normalization layer, and an LReLU activation layer are used to build the encoder. The decoder consists of a deconvolution layer, a normalization layer, and an LReLU activation layer. The structure of the generator $G$ is U-shaped [32] and is composed of $n$ encoders and decoders along with the SE attention module detailed in the next subsection. The SE module is tasked with ignoring the less important features conveyed from the convolution operation in the decoder and the encoder while enhancing the important ones. The reweighted features obtained from the SE module are then fed to the subsequent layers.

The discriminator $D$ consists of a convolution layer, an LReLU activation layer, and three encoders, followed by another convolution layer and a Tanh layer. Given a green-tide image $I$, the generator generates a green-tide detection image $S$. The image $I$ and the generated detection image $S$ or the reference image $R$ are regarded as the inputs of the discriminator. The discriminator generates a representation of the variational variable defined as $v$ [33] for the image. Adversarial training of the generator and the discriminator seeks to minimize $f$-divergence so that the generated detection images are close to the reference images.

### 2) THE SE MODULE

The output from the convolution operation is generated by summing all channels, and the relationship between the channels is invisibly entangled in the spatial information. Hence, the purpose of introducing the SE module is to improve network sensitivity to the important features by modeling the interdependence of channels. As a consequence, less useful features are suppressed. The SE module forms part of a U-shaped structure in the generator immediately following the encoders and decoders [34]. The module comprises two separate operations – squeeze and excitation – as proposed in [35]. The output obtained through the convolution operation of the decoder and the encoder $U$ contains spatial and channel information. $U$ is regarded as the input to the SE module for feature reconstruction. Here, we assume the input to the SE module $U = [u_1, u_2, \cdots, u_C]$ is the combination of channels $u_i \in \mathrm{R}^{H \times W}$. The feature maps $U$ from the convolution operation are then squeezed into a channel descriptor through a global average pooling operation [36]. $sq \in \mathrm{R}^{1\times1\times C}$ is produced by squeezing $U$ with spatial dimensions $H \times W$. The $k^{th}$ element of $sq \in \mathrm{R}^{1\times1\times C}$ is expressed as follows:

$$sq_k = F_{sq}(u_k) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_k(i,j). \tag{2}$$

Thus, the squeeze operation embeds the global spatial information into the vector $sq$, which expresses all the information in the image.

For the purposes of better using the integrated global information obtained through the squeeze operation, the excitation operation comprehensively derives the channel-wise dependencies. The output $sq$ from the global average pooling operation is transformed to $W_2(\delta(W_1 sq))$. Here, $W_1 \in \mathrm{R}^{C/r\times C}$ and $W_2 \in \mathrm{R}^{C\times C/r}$ are the weights of two fully connected layers [37] in which $r$ is set to 2. A bottleneck between the two fully connected layers is then formed around the nonlinearity in the excitation block. In more detail, the structure of the excitation operation is a dimensionality-reduction layer with $W_1$, a ReLU layer, and a dimensionality-increasing layer with $W_2$ as illustrated in Fig. 4. A sigmoid activation function

$\sigma(\cdot)$ constrains $W_2(\delta(W_1 sq))$ to the interval [0,1]. A feature vector with a channel dimension of $C/r$ is constructed through the first fully connected layer and the ReLU layer. This feature vector is then converted into a vector with a channel dimension of $C$ by the second fully connected layer and the sigmoid layers. The output of the last sigmoid layer is therefore:

$$ex = F_{ex}(sq, W) = \sigma(W_2 \delta(W_1 sq)). \tag{3}$$

where $ex$ indicates the importance of the $i^{th}$ channel. The final output is obtained by scaling $U$ with $ex$:

$$\tilde{u}_k = F_{scale}(u_k, ex_k) = ex_k . u_k. \tag{4}$$

In the excitation operation, reducing the dimensionality reduction of the first fully connected layer is designed to reduce the computational complexity of the network, while the increasing the dimensionality with the second fully connected layer ensures that the dimensions are the same as that of the original $U$ for calculating the weight of each layer feature. $U$ is recalibrated into the form of $U = [ex_1 u_1, ex_2 u_2, \ldots, ex_C u_C]$. The SE attention module ignores the less important channels and emphasizes the important channels adaptively. The process of the SE module is shown in Fig. 4, and where it sits in the architecture of the generator is shown in Fig. 5.

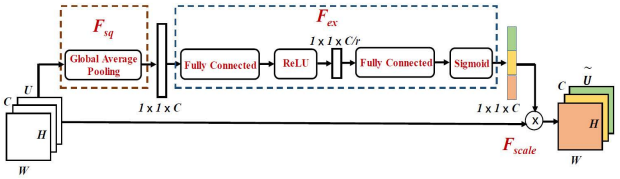



**FIGURE 4.** The SE module architecture. The SE attention module contains a squeeze block and an excitation block. The squeeze block is performed via global average pooling. The excitation is composed of a bottleneck of two fully connected layers around a non-linearity Ultimately, the SE module reweights feature importance.

### 3) THE TRAINING SCHEME

The overall objective of the training scheme is to minimize the $f$-divergence between the generated image and the reference image. When properly trained, the generator should be able to generate accurate detection images that are very similar to the reference images. In statistics, $f$-divergence is a function used to measure the difference between both probability distributions – that is, how much both distributions are the same or different. The distributions of the reference image $R$ and the generated detection image $S$ are denoted as $P_R$ and $P_S$, respectively. The $f$-divergence of $P_R$ and $P_S$ is defined as follows:

$$D_f(P_R||P_S) = \int f\left(\frac{P_R(x)}{P_S(x)}\right) P_S(x)\, dx. \tag{5}$$

where $x$ is the distribution variable, and $f(\cdot)$ is a function representing the distribution divergence. $p_R(x)$ and $p_S(x)$ are density functions of the reference detection distribution $P_R$ and the generated detection distribution $P_S$, respectively.

Considering that directly minimizing $f$-divergence is difficult, SE-GAN minimizes the tight lower bound of the $f$-divergence instead, and a surrogate loss function $\phi(\cdot)$ [38] is introduced to reformulate $f$-divergence. $\phi(\cdot)$ is formulated as:

$$f\left(\frac{p_R(x)}{p_S(x)}\right) = -\inf_v \left[\phi(-v) + \phi(v)\frac{p_R(x)}{p_S(x)}\right]. \tag{6}$$

The $f$-divergence between the reference image $R$ and the generated detection image $S$ is then reformulated as:

$$\begin{aligned} D_f(P_R||P_S) &= \int f\left(\frac{P_R(x)}{P_S(x)}\right) P_S(x)\, dx \\ &= \int \left\{-\inf_v[\phi(-v) + \phi(v)\frac{p_R(x)}{p_S(x)}\right\} P_S(x)\, dx \\ &\geq -\inf_v [\int \phi(-v) P_S(x)\, dx + \int \phi(v) p_R(x)\, dx] \\ &= -\inf_v \mathrm{E}_{x\sim P_R}[\phi(v)] + \mathrm{E}_{x\sim P_S}[\phi(-v)] \\ &= \max_v \mathrm{E}_{x\sim P_R}[\phi(v)] + \mathrm{E}_{x\sim P_S}[\phi(-v)]. \end{aligned} \tag{7}$$

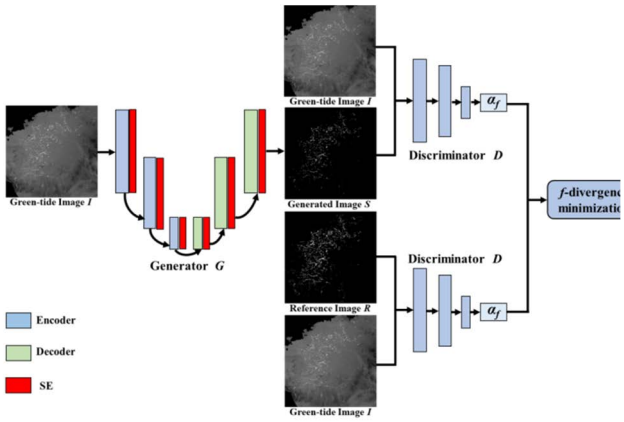



**FIGURE 5.** The process of minimizing $f$-divergence. The generator is a U-shaped structure composed of several encoders, decoders and SE modules while the discriminator is composed of encoders. Given a green-tide image $I$ in training, the generator generates an accurate detection image $S$. The green-tide image $I$ and the generated detection image $S$ or the reference image $R$ are regarded as the inputs to the discriminator. The $f$-divergence is minimized by adversarially training the generator and the discriminator.

The relationship of the variational variable $v$ and the distribution variable $x$ of the generated detection image or the reference image is built through the discriminator. The inequality mentioned above follows the inequality in [39]. The lower bound of the $f$-divergence is expressed through the inequality and is more constrained due to the maximization in (8). As mentioned, the $f$-divergence is minimized by minimizing the tight lower bound of the $f$-divergence as follows:

$$\min_{p_S} \max_v \mathrm{E}_{x\sim P_R}[\phi(v)] + \mathrm{E}_{x\sim P_S}[\phi(-v)]. \tag{8}$$

Owing to this $max - min$ operation, SE-GAN does minimize the difference between the generated detection images

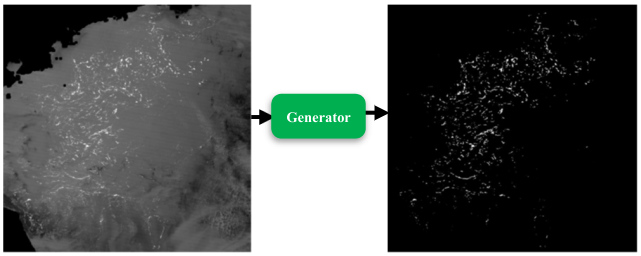
and the reference images. The training scheme uses a transformation function $\alpha_f(\cdot)$ to scale the variational variable $v$ obtained from the discriminator. The combination of $\phi(\cdot)$ and $\alpha_f(\cdot)$ means the $f$-divergence is varied and generalized. In this paper, we have used just one form of $f$-divergence, that is, Pearson's $\chi^2$ divergence. $\theta_g$ and $\theta_d$ represent the parameters of the generator $G$ and the discriminator $D$, respectively. The final optimization objective transformed from (9) is therefore:

$$\min_{\theta_g}\max_{\theta_d,\theta_g} \mathrm{E}_{R\sim P_R, I\sim P_I}\left[\phi\left(\alpha_f\left(D\left(R,I\right)\right)\right)\right] + \mathrm{E}_{I\sim P_I}\left[\phi\left(-\alpha_f\left(D\left(G\left(I\right),I\right)\right)\right)\right] \quad (9)$$

The $max-min$ optimization objective reveals the adversarial relationship between the generator and the discriminator. In the maximization operation, the discriminator's task is to distinguish the generated detection image from the reference image to the greatest extent possible. In the minimization operation, the generator's task is to generate a detection image that will fool the discriminator into predicting that the image is a reference image. The maximization operation achieves the maximal lower bound of the $f$-divergence, while the minimization operation achieves the minimal value of the maximal lower bound. It is the adversarial training of the generator and the discriminator that achieves the minimum of the maximal lower bound of the $f$-divergence. Given a green-tide image $I$ in the training, the generator generates an accurate detection image $S$. The green-tide image $I$ and the generated detection image $S$ or the reference image $R$ are regarded as the inputs of the discriminator. The $f$-divergence is minimized by adversarial training the generator and the discriminator as shown in Fig. 5. When the adversarial training is completed, the abilities of both the generator and the discriminator are strengthened so that the generator produces a green-tide detection image very similar to the reference image. As a result, when an arbitrary green-tide image is input into SE-GAN in practical situations, the generator will generate an accurate green-tide detection image, as shown in Fig. 6.

## IV. EXPERIMENT AND RESULTS

In view of the largest distribution of green-tides in the Yellow Sea, the MODIS images covering all monitored sea areas often disturbed diverse seawater transparency and changing marine environments. Similar machine learning algorithms tend to focus more on the green-tide monitoring of small-format images, and do not test or verify the monitoring efficiency of large-format images. The detection results for large-format green tide images either have poor detection efficiency or poor consistency of results. Therefore, we designed the detection experiments of large-format and small-format images in the algorithm testing process. The results show that SE-GAN provides excellent detection for both types of images. The experimental details are as follows.




(a) Mask Image (b) Separation Image

**FIGURE 6.** The inference process. In practical operations, the generator automatically generates an accurate detection result for any given green-tide image.

### A. DATA AND TRAINING SET

2019 was the year with the largest distribution area and longest duration of green-tides in the Yellow Sea since 2010. As described in II-B, we selected about 90 valid MODIS images from May to August 2019 to construct our experiment dataset. All experiment images were pre-processed with geometric correction, atmospheric correction. Only 20 images were selected from the experiment dataset to construct the training dataset. All images were dealt as follows. First, we used the artificial interpretation method to manually adjust and set the appropriate threshold to extract the green-tide while generating the corresponding binary images as the reference images. In these reference images, the black pixels corresponded to the background, and the white pixels corresponded to the green-tides. Second, we put these reference images with image pairs pre-processed by NDVI calculating and sea-land separation filters, and fixed the learning rate as $\eta = 0.0002$, and set the minibatch as 1 [31].

### B. EVALUATION CRITERIA

To quantitatively evaluate SE-GAN's effectiveness we used precision, recall and $F1$-score as the evaluation criteria. The assessment worked as follows. Let $i$ represent the pixels of the generated detection image $S$ and the reference image $R$. Hence, $S^{(i)}$ denotes the pixels of the generated detection image $S$, and $R^{(i)}$ denotes the pixels of the reference image $R$. If $S^{(i)} + R^{(i)} = 2$, the pair was deemed a true positive ($TP$). Likewise, a false positive (FP) was assigned from $S^{(i)} - R^{(i)} = 1$; $S^{(i)} + R^{(i)} = 0$ represented a true negative ($TN$); and $S^{(i)} - R^{(i)} = -1$ a false negative ($FN$). The evaluation criteria were then are calculated by:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (10)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (11)$$

$$F1\text{-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

### C. COMPARISON METHODS

We compared SE-GAN to the threshold method [7], the HSV color space method [8], DeepLabv3+ [40], FCN [41]

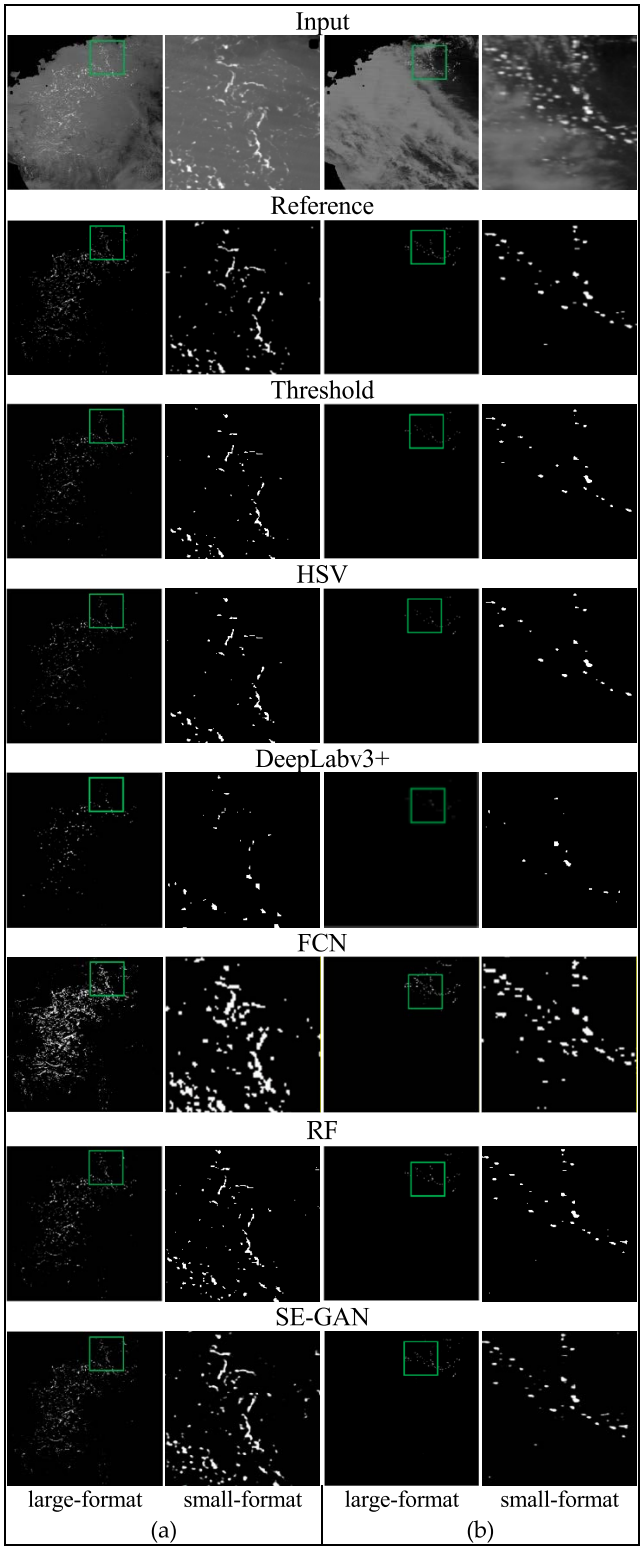



**FIGURE 7.** Comparison of the detection results obtained from six different methods and the reference image. Image (a) shows the cloudless scene. Image (b) shows the cloudy scene.

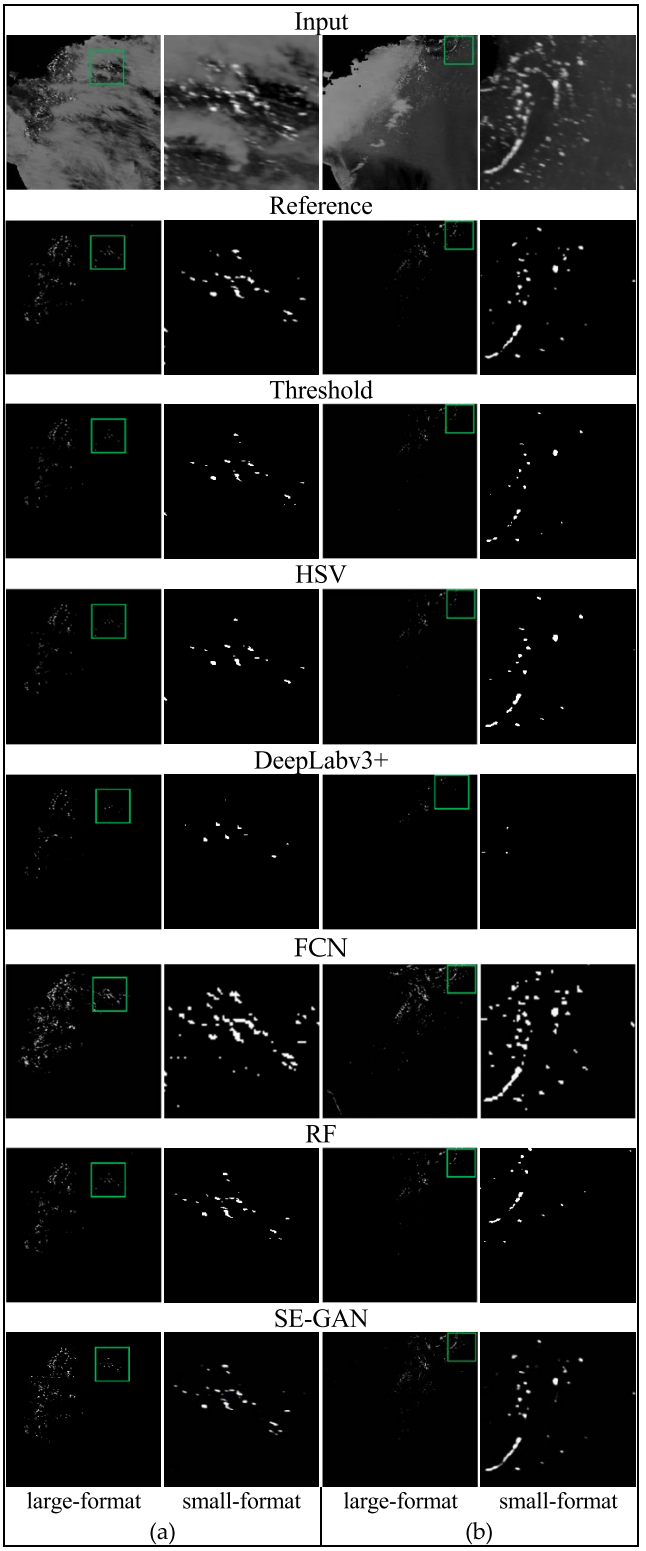



**FIGURE 8.** Comparison of the detection results obtained from six different methods and the reference image. Image (a) and image (b) all show the cloud coverage scenarios. Image (a) shows green-tide monitoring in cloud crevices. Image (b) shows the result in clouds of gradient thickness.

and random forest (RF) classification [20]. The threshold method requires manual adjustment of the appropriate threshold which is affected by subjective factors. Each image requires a specific threshold or several thresholds for different areas, excessive artificial intervention leads to lower detection efficiency. The HSV method has three threshold constraints

**TABLE 1. Performance of large-format green-tide detection.**

| Method | Criterion | Fig. 7(a) | Fig. 7(b) | Fig.8(a) | Fig. 8(b) |
|---|---|---|---|---|---|
| Threshold | Recall | 51.49% | 48.44% | 51.31% | 40.54% |
| | Precision | 99.96% | 100.00% | 99.83% | 100.00% |
| | F1-score | 67.97% | 65.26% | 67.78% | 57.69% |
| HSV | Recall | 55.51% | 50.35% | 52.06% | 48.07% |
| | Precision | 99.98% | 100.00% | 99.79% | 100.00% |
| | F1-score | 71.39% | 66.98% | 68.42% | 64.93% |
| SSS | Recall | 25.51% | 22.71% | 29.12% | 17.45% |
| | Precision | 77.87% | 85.93% | 85.98% | 87.50% |
| | F1-score | 38.43% | 35.93% | 43.51% | 29.10% |
| FCN | Recall | 100.00% | 86.92% | 98.51% | 89.60% |
| | Precision | 38.45% | 32.13% | 46.78% | 22.93% |
| | F1-score | 55.54% | 46.92% | 63.44% | 36.52% |
| RF | Recall | 76.24% | 74.67% | 76.33% | 66.89% |
| | Precision | 99.87% | 99.87% | 99.64% | 98.84% |
| | F1-score | 86.47% | 85.45% | 86.44% | 79.78% |
| SE-GAN | Recall | 98.19% | 91.26% | 91.92% | 92.43% |
| | Precision | 57.31% | 88.59% | 82.21% | 85.03% |
| | F1-score | 72.38% | 89.90% | 86.80% | 88.58% |

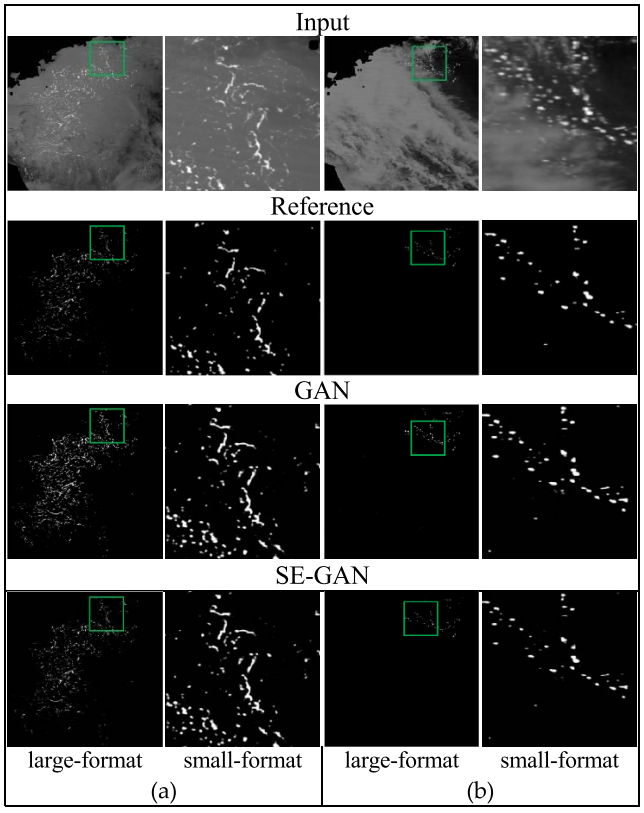



**FIGURE 9. Comparison of the detection results obtained from GAN without and with SE module based on the images in Fig.7.**

of H, S, and V, which makes the detection results of the HSV method have higher reliability than the detection results of the threshold method. However, the ranges of H, S, and V are manually adjusted according to the matching degree between

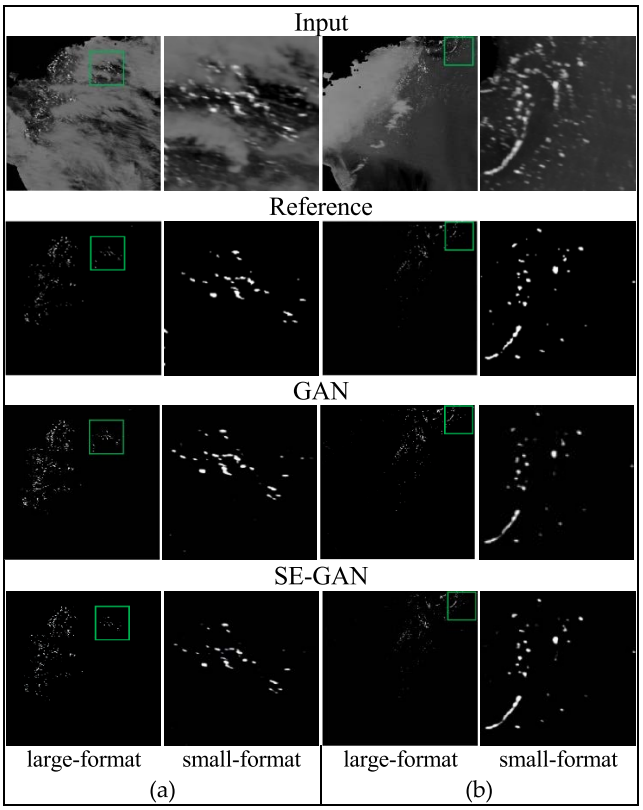



**FIGURE 10. Comparison of the detection results obtained from GAN without and with SE module based on the images in Fig.8.**

**TABLE 2. Performance of gan on large-format green tide images without or with SE module.**

| Method | Criterion | Fig. 7(a) | Fig. 7(b) | Fig. 8(a) | Fig. 8(b) |
|---|---|---|---|---|---|
| GAN | Recall | 98.65% | 100.00% | 98.05% | 98.78% |
| | Precision | 54.82% | 60.71% | 68.22% | 64.20% |
| | F1-score | 70.47% | 75.55% | 80.46% | 77.82% |
| SE-GAN | Recall | 98.19% | 91.26% | 91.92% | 92.43% |
| | Precision | 57.31% | 88.59% | 82.21% | 85.03% |
| | F1-score | 72.38% | 89.90% | 86.80% | 88.58% |

the detection results and the green-tide monitoring images, and too much manual intervention is also undesirable. The DeepLabv3+ method used in this paper uses DeepLabv3 as the encoder architecture, on which the decoder module is added to refine the segmentation results. Its backbone network is a deep convolutional neural network (DCNN) with atrous convolution in the encoder, and the low-level and high-level features are fused in the decoder to obtain detection results. Here, we modified downsample_factor as 16 and got relatively good results. The FCN method uses the convolution layer to replace the full connection layer of CNN. It was a milestone in the semantic segmentation and always ranked first in application for consecutive years. The RF method is one of the most commonly used and powerful supervised learning algorithms. It is an algorithm that integrates multiple

**TABLE 3. Average performance of the six detection methods on thirty large-format green tide images.**

| Method | Recall | Precision | F1-score |
|---|---|---|---|
| Threshold | 44.56% | 99.35% | 61.53% |
| HSV | 48.92% | 98.99% | 65.46.% |
| DeepLabv3+ | 19.76% | 87.15% | 32.22% |
| FCN | 98.02% | 46.91% | 63.45% |
| RF | 63.42% | 95.17% | 69.87% |
| SE-GAN | 76.97% | 88.87% | 82.49% |

**TABLE 4. Training and testing costs of different methods on large-format green tide images.**

| Method | Training time | Testing time |
|---|---|---|
| DeepLabv3+ | 1.5hours | 10seconds |
| FCN | 2hours | 10minutes |
| RF | 20 minutes | 109hours |
| SE-GAN | 11 hours 44 minutes | 2 minutes |

decision trees. The output category is determined by the mode of individual number output. Although the RF method is relatively accurate, it requires several times of the training and testing cycles of other methods, resulting in extremely low detection efficiency. In the RF method, we set the parameters including n_estimators, max_depth, random_state of RandomForestClassifier as 250, 12, 42 respectively, which is suitable to achieve good detection results.

### 1) LARGE-FORMAT RESULTS

To qualitatively evaluate our method, we chose four typical scenarios of large-format MODIS images with a resolution of 1024 × 1024. Images in Fig. 7 and Fig. 8 represent four typical green-tide monitoring scenarios. Fig. 7 (a) contains more green-tide and less clouds. Fig. 7 (b) shows most of green-tides covered by thick clouds, and few green-tides at thick cloud boundaries are needed to be detected. Fig. 8 (a) shows the monitoring area are covered by thick and thin clouds, and green-tide mixes with them. Fig. 8(b) shows green-tide with thin clouds. The four green-tide images are acquired by Terra satellite on June 24, June 25, June 27, and July 6, 2019, respectively. The detection results of five methods are shown in both figures. The top row of each figure shows the input green-tide image, followed by the reference image, then the results obtained from the threshold method, the HSV method, the DeepLabv3+ method, the FCN method, the RF method, and our SE-GAN method. Each figure has two columns, the left column is the large-format detection results of our study area, right column shows a small-format image representing the close-up of the green box in the left image. It is clear from visual inspection that SE-GAN's results are more complete than those obtained from the other detection methods in terms of the overall distribution and the density of green-tide. From the detailed detection results in the right column, it is also clear that the

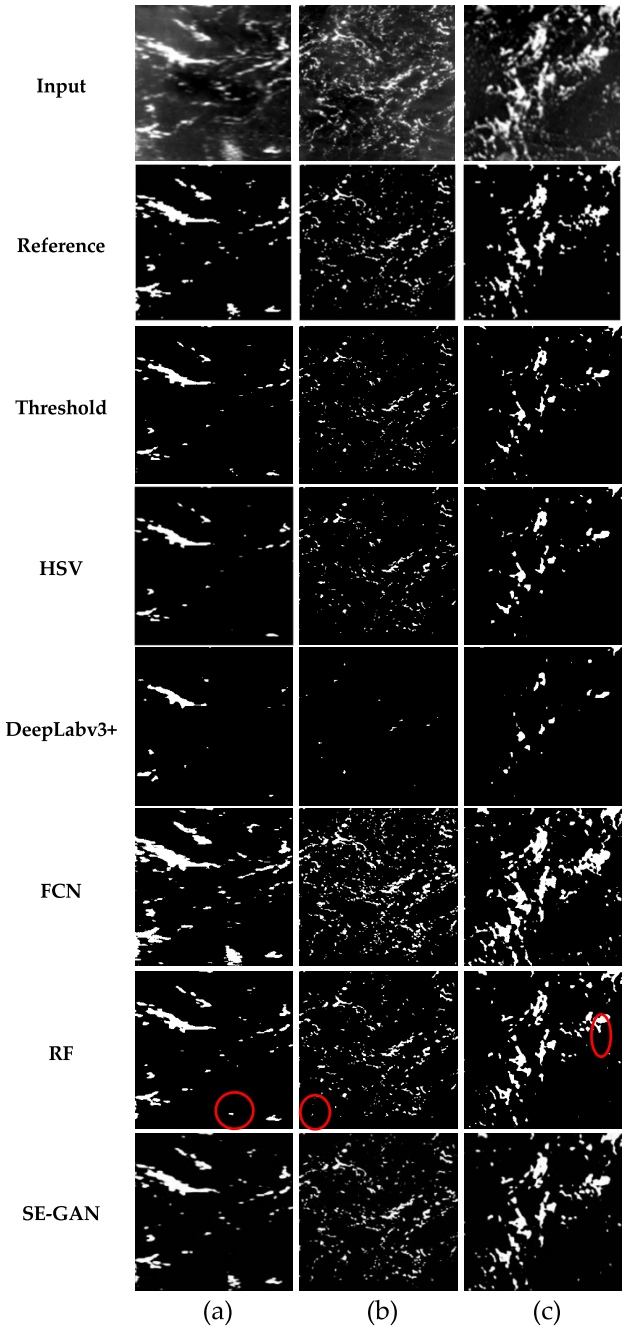



**FIGURE 11. Comparison of the detection results obtained from six different methods and the reference image.**

results from SE-GAN are clearer with almost no green-tide missed. Moreover, SE-GAN was able to produce a map that was much closer to the reference image than the other methods. Considering the amount of cloud interference, especially in images shown in Fig. 7 (b) and Fig. 8 (a)(b), this is a very promising result in terms of accuracy and adaptability.

Table 1 lists the quantitative evaluations for recall, precision and $F1$-score of these four large-format images. As the table shows, the recall values of SE-GAN were higher than that of all other methods, while almost all methods posted higher precision values than SE-GAN. This is because they

**TABLE 5.** Performance of small-format green-tide detection.

| Method | Criterion | (a) | (b) | (c) |
|---|---|---|---|---|
| Threshold | Recall | 50.37% | 46.69% | 42.27% |
| | Precision | 100.00% | 100.00% | 100.00% |
| | F1-score | 67.00% | 63.65% | 59.43% |
| HSV | Recall | 51.43% | 48.30% | 43.62% |
| | Precision | 100.00% | 100.00% | 99.87% |
| | F1-score | 67.93% | 65.14% | 60.72% |
| DeepLabv3+ | Recall | 45.47% | 6.70% | 40.86% |
| | Precision | 100.00% | 100.00% | 100.00% |
| | F1-score | 62.51% | 12.55% | 58.02% |
| FCN | Recall | 98.01% | 96.99% | 96.98% |
| | Precision | 63.01% | 64.64% | 64.28% |
| | F1-score | 76.71% | 77.58% | 77.31% |
| RF | Recall | 73.96% | 75.86% | 79.95% |
| | Precision | 98.55% | 99.55% | 94.85% |
| | F1-score | 84.50% | 86.11% | 86.76% |
| SE-GAN | Recall | 92.12% | 96.20% | 91.38% |
| | Precision | 92.29% | 86.92% | 86.67% |
| | F1-score | 92.20% | 91.33% | 88.96% |

detected less green-tide, although what they detected was accurate. The $F1$-score is a comprehensive index combining recall and precision. Here, SE-GAN method had the highest $F1$-scores for the three images with more cloud. Although RF has the highest precision values, it took several magnitudes longer to achieve a comparable result.

To validate the effectiveness of SE module in the feature extraction, based on the 4 images in Fig.7 and Fig.8, the comparison experiments have been carried out, and the results are shown in Fig.9 and Fig.10. Based on the statistical results in Table 2, SE-GAN has the higher $F1$-score than GAN.

In addition to the 4 images shown in Fig. 7 and Fig. 8, we randomly selected 26 other images and combined them to form 30 images for statistical analysis of six methods. Average recall, precision, and F1-scores for the eight large-format images are listed in Table 3. Again, SE-GAN method yielded the highest values in terms of F1-score.

In terms of training 20 large-format green-tide images and testing 30 large-format green-tide images for all methods, the time cost results are shown in Table 4. Although the SE-GAN method needs the most time in training, its testing time is less.

### 2) SMALL-FORMAT RESULTS

To test SE-GAN with small-format images, we chose three MODIS images with less cloud and cropped to a size of 1/16 of large-format image, i.e., 256 × 256 pixels. Fig. 11 shows the results of the three images. Here again, SE-GAN's results look clearer and more comprehensive than those of the other five methods. In particular, for green-tide that was obviously ignored or detected wrong by the RF algorithm, we marked the red circles in RF results.

**TABLE 6.** Average performance of the six detection methods on twenty-seven small-format green tide images.

| Method | Recall | Precision | F1-score |
|---|---|---|---|
| Threshold | 52.38% | 98.98% | 68.51% |
| HSV | 56.78% | 98.75% | 72.10% |
| DeepLabv3+ | 24.56% | 99.00% | 39.36% |
| FCN | 97.96% | 52.13% | 68.05% |
| RF | 75.10% | 94.36% | 82.65% |
| SE-GAN | 92.26% | 87.70% | 89.77% |

Table 5 lists the recall, precision and $F1$-scores of the results of three images shown in Fig. 11. Same as the large-format results, for SE-GAN, $F1$-scores of three small-format images are 92.20%, 91.33% and 88.96% respectively, which are all the highest compared with the other five methods and represent the highest performance. But compared with Table 1, recalls of SE-GAN are 92.12%, 96.20% and 91.38%, and precisions of SE-GAN are 92.29%, 86.92% and 86.67% respectively, which has significantly better performance.

Again, in addition to the 3 images shown in Fig. 11, we randomly selected 24 images and combined them to form 27 images for statistical analysis of six methods. Table 6 lists the average performance over the 27 images. Same as with the large-format images, SE-GAN outperformed the other methods in terms of $F1$-score with the small-format images.

## V. CONCLUSION

This paper presented a novel framework for detecting green-tide from MODIS images called SE-GAN. The framework begins by parsing the images through two preprocessing filters: one that applies NDVI values and another that separates land from sea. This enhances the images and removes interference from clouds and land masses. The architecture comprises a generator and a discriminator, where the generator consists of encoders, decoders, and a "squeeze and excitation" attention module, all configured in a U-shaped structure. The training scheme is adversarial with the objective of minimizing $f$-divergence. Once built, the model can automatically detect green-tide in any MODIS image. Experiments confirm that SE-GAN achieves more accurate detection results with both large and small format images than four other traditional detection methods. Compared with other previous methods, our proposed method is more efficient, more accurate, and requires less data.

According to the characteristics of the green-tide in the Yellow Sea, we designed the fine-grained monitoring experiments in both large and small scenarios, which could comprehensively verify the sensitivity and accuracy of green-tide extraction. The large-format can verify the accuracy of the identification of the distribution range of green-tide, and the small-format can test the extraction amount of green-tide distributing in the nearshore and offshore areas, which meets the application requirements of the emergency department.

SE-GAN algorithm's training only needs 20 pairs of images and it is automatic without interaction, which makes the model be flexibly optimized and adjusted easily when needed. It is easy to operate and very suitable for the operational green-tide monitoring with inter-annual variability.

## REFERENCES

[1] I. Valiela, J. Mcclelland, J. Hauxwell, and P. J. Behr, "Macroalgal Blooms in shallow estuaries: Controls and ecophysiological and ecosystem consequences," *Limnol. Oceanogr.*, vol. 42, no. 5, pp. 1105–1118, Jul. 1997, doi: 10.4319/lo.1997.42.5_part_2.1105.

[2] Z. Wang, J. Xiao, S. Fan, Y. Li, X. Liu, and D. Liu, "Who made the world's largest green tide in China?—An integrated study on the initiation and early development of the green tide in Yellow Sea," *Limnol. Oceanogr.*, vol. 60, no. 4, pp. 1105–1117, Apr. 2015, doi: 10.1002/lno.10083.

[3] J. Zhang, J. Shi, S. Gao, and Y. Huo, "Annual patterns of macroalgal Blooms in the Yellow Sea during 2007–2017," *PLoS ONE*, vol. 14, no. 1, pp. 1–12, Jan. 2019, doi: 10.1371/journal.pone.0210460.

[4] Y. Zhang, P. He, H. Li, G. Li, J. Liu, F. Jiao, J. Zhang, Y. Huo, X. Shi, R. Su, N. Ye, D. Liu, R. Yu, Z. Wang, M. Zhou, and N. Jiao, "*Ulva prolifera* green-tide outbreaks and their environmental impact in the Yellow Sea, China," *Nat. Sci. Rev.*, vol. 6, no. 4, pp. 825–838, Jul. 2019, doi: 10.1093/nsr/nwz026.

[5] L. Schreyers, T. van Emmerik, L. Biermann, and Y.-F. Le Lay, "Spotting green tides over Brittany from space: Three decades of monitoring with landsat imagery," *Remote Sens.*, vol. 13, no. 8, p. 1408, Apr. 2021, doi: 10.3390/rs13081408.

[6] B. Wang and L. Wu, "Numerical study on the massive outbreak of the *Ulva prolifera* green tides in the southwestern Yellow Sea in 2021," *J. Mar. Sci. Eng.*, vol. 9, no. 11, p. 1167, Oct. 2021, doi: 10.3390/jmse9111167.

[7] X. Gu, X.-F. Chen, Q. Yin, Z.-Q. Li, H. Xu, Y. Shao, and Z.-W. Li, "Stereoscopic remote sensing used in monitoring Enteromorpha Prolifra disaster in Chinese Yellow Sea," *Spectrosc. Spectral Anal.*, vol. 31, no. 6, pp. 1627–1632, Jun. 2011, doi: 10.3964/j.issn.1000-0593(2011)06-1627-06.

[8] E. Prasetyo, R. D. Adityo, N. Suciati, and C. Fatichah, "Mango leaf image segmentation on HSV and YCbCr color spaces using Otsu thresholding," in *Proc. 3rd Int. Conf. Sci. Technol.-Comput. (ICST)*, Jul. 2017, pp. 99–103, doi: 10.1109/ICSTC.2017.8011860.

[9] N. Wang, J. Huang, and T. Cui, "Capability comparison of 5 vegetation indices for detecting the green tide in different development phases and the application," *Acta Laser Biol. Sinica*, vol. 23, no. 6, pp. 590–595, 2014.

[10] Q. Xing, L. Wu, L. Tian, T. Cui, L. Li, F. Kong, X. Gao, and M. Wu, "Remote sensing of early-stage green tide in the Yellow Sea for floating-macroalgae collecting campaign," *Mar. Pollut. Bull.*, vol. 133, pp. 150–156, Aug. 2018, doi: 10.1016/j.marpolbul.2018.05.035.

[11] X. Jin, J. Wan, W. Hu, Y. Song, and B. Lu, "Retrieval of green tide concentration and interannual variation analysis in Yellow Sea based on multi-source remote sensing monitoring," in *Proc. Global Oceans: Singapore-US Gulf Coast*, Oct. 2020, pp. 1–5, doi: 10.1109/IEEECONF38699.2020.9388992.

[12] T.-W. Cui, J. Zhang, L.-E. Sun, Y.-J. Jia, W. Zhao, Z.-L. Wang, and J.-M. Meng, "Satellite monitoring of massive green macroalgae Bloom (GMB): Imaging ability comparison of multi-source data and drifting velocity estimation," *Int. J. Remote Sens.*, vol. 33, no. 17, pp. 5513–5527, Sep. 2012, doi: 10.1080/01431161.2012.663112.

[13] L. Qi, C. Hu, Q. Xing, and S. Shang, "Long-term trend of *Ulva prolifera* Blooms in the western Yellow Sea," *Harmful Algae*, vol. 58, pp. 35–44, Sep. 2016.

[14] C. Hu, "A novel ocean color index to detect floating algae in the global oceans," *Remote Sens. Environ.*, vol. 113, pp. 2118–2129, Oct. 2009, doi: 10.1016/j.rse.2009.05.012.

[15] W. Shi and M. Wang, "Green macroalgae Blooms in the Yellow Sea during the spring and summer of 2008," *J. Geophys. Res.*, vol. 114, no. C12, pp. 1–10, Dec. 2009, doi: 10.1029/2009JC005513.

[16] X. Dong, J. Dong, and L. Qu, "Enteromorpha detection in aerial images using support vector machines," in *Proc. IEEE Youth Conf. Inf., Comput. Telecommun.*, Sep. 2009, pp. 299–302, doi: 10.1109/YCICT.2009.5382365.

[17] S. Wu, F. Shao, Y. Wang, R. Sun, and J. Wang, "Enteromorpha prolifera detection with MODIS image using semi-supervised clustering," *J. Comput.*, vol. 9, no. 5, pp. 1259–1265, May 2014, doi: 10.4304/jcp.9.5.1259-1265.

[18] A. Dogliotti, J. Gossn, Q. Vanhellemont, and K. Ruddick, "Detecting and quantifying a massive invasion of floating aquatic plants in the Río de la Plata turbid waters using high spatial resolution ocean color imagery," *Remote Sens.*, vol. 10, no. 7, p. 1140, Jul. 2018, doi: 10.3390/rs10071140.

[19] L. Qu and X. Dong, "Enteromorpha prolifera detection in aerial images based on image retrieval," in *Proc. 3rd Int. Conf. Knowl. Discovery Data Mining*, Jan. 2010, pp. 25–28, doi: 10.1109/WKDD.2010.144.

[20] C. Xie, J. Dong, F. Sun, and L. Bing, "Object-oriented random forest classification for Enteromorpha prolifera detection with SAR images," in *Proc. Int. Conf. Virtual Reality Visualizat. (ICVRV)*, Sep. 2016, pp. 119–125, doi: 10.1109/ICVRV.2016.27.

[21] S. Wang, L. Liu, L. Duan, C. Yu, G. Cai, F. Gao, and J. Dong, "Accurate segmentation of *Ulva prolifera* regions with superpixel and CNNs," in *Proc. Int. Conf. Secur., Pattern Anal., Cybern. (SPAC)*, Dec. 2017, pp. 433–438, doi: 10.1109/SPAC.2017.8304318.

[22] H. Yin, Y. Liu, and Q. Chen, "An elegant end-to-end fully convolutional network (E3FCN) for green tide detection using MODIS data," in *Proc. 10th IAPR Workshop Pattern Recognit. Remote Sens. (PRRS)*, Beijing, China, Aug. 2018, pp. 1–6.

[23] Z. Wang, Z. Fang, Y. Wu, J. Liang, and X. Song, "Multi-source evidence data fusion approach to detect daily distribution and coverage of *Ulva prolifera* in the Yellow Sea, China," *IEEE Access*, vol. 7, pp. 115214–115228, 2019, doi: 10.1109/ACCESS.2019.2936247.

[24] Y. Cao, Y. Wu, Z. Fang, X. Cui, J. Liang, and X. Song, "Spatiotemporal patterns and morphological characteristics of *Ulva prolifera* distribution in the Yellow Sea, China in 2016–2018," *Remote Sens.*, vol. 11, no. 4, p. 445, Feb. 2019, doi: 10.3390/rs11040445.

[25] V. I. Haltrin, "Chlorophyll-based model of seawater optical properties," *Appl. Opt.*, vol. 38, no. 33, pp. 6826–6832, Dec. 1999, doi: 10.1364/AO.38.006826.

[26] Q. Xing, C. Hu, D. Tang, L. Tian, S. Tang, X. Wang, M. Lou, and X. Gao, "World's largest macroalgal Blooms altered phytoplankton biomass in summer in the Yellow Sea: Satellite observations," *Remote Sens.*, vol. 7, no. 9, pp. 12297–12313, Sep. 2015, doi: 10.3390/rs70912297.

[27] M. El-Gammal, R. Ali, and R. Samra, "NDVI threshold classification for detecting vegetation cover in Damietta governorate, Egypt," *J. Amer. Sci.*, vol. 10, no. 8, pp. 108–113, Jan. 2014.

[28] K. B. Larson and A. R. Tuor, "Deep learning classification of cheatgrass invasion in the western United States using biophysical and remote sensing data," *Remote Sens.*, vol. 13, no. 7, p. 1246, Mar. 2021, doi: 10.3390/rs13071246.

[29] T. Karin. *Global Land Mask*. Accessed: Oct. 5, 2020. [Online]. Available: https://doi.org/10.5281/zenodo.4066722

[30] I. Goodfellow, J. Pouget-Abadie, and M. Mirza, "Generative adversarial networks," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, Montreal, QC, Canada, Dec. 2014.

[31] H. Liu, B. Jiang, and Y. Song, "Rethinking image inpainting via a mutual encoder-decoder with feature equalizations," in *Proc. Comput. Vis.-ECCVm 16th Eur. Conf.*, Glasgow, U.K., Aug. 2020, pp. 23–28.

[32] E. Schonfeld, B. Schiele, and A. Khoreva, "A U-Net based discriminator for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2020, pp. 15–20.

[33] X. Yu, H. Zhang, C. Luo, H. Qi, and P. Ren, "Oil spill segmentation via adversarial $f$-divergence learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, pp. 4973–4988, 2018.

[34] A. G. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel squeeze & excitation in fully convolutional networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2018, pp. 1803–1811.

[35] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, CA, USA, Jun. 2018, pp. 18–23.

[36] A. Al-Sabaawi, H. M. Ibrahim, Z. M. Arkah, M. Al-Amidie, and L. Alzubaidi, "Amended convolutional neural network with global average pooling for image classification," in Proc. Int. Conf. Intell. Syst. Design Appl., Xian, China, Dec. 2020, pp. 11–13.

[37] S. H. S. Basha, S. R. Dubey, V. Pulabaigari, and S. Mukherjee, "Impact of fully connected layers on performance of convolutional neural networks for image classification," *Neurocomputing*, vol. 378, pp. 112–119, Feb. 2020.

[38] X. Nguyen, M. J. Wainwright, and M. I. Jordan, "On surrogate loss functions and $f$-divergences," *Ann. Statist.*, vol. 37, no. 2, pp. 876–904, Apr. 2009.
[39] D. Chandler, *Introduction to Modern Statistical Mechanics*. Oxford, U.K.: Oxford Univ. Press, 1987, p. 40.
[40] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, Sep. 2018, pp. 8–14.
[41] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017, doi: 10.1109/CVPR.2015.7298965.

**XIFANG JIN** received the M.E. degree in software engineering from Zhejiang University, Hangzhou, China, in 2006. She is currently pursuing the Ph.D. degree in computer technology and resource information engineering with the China University of Petroleum (East China), Qingdao, China. She joined the North China Sea Marine Forecasting Center, Ministry of Natural Resources (MNR), Qingdao, and was a Professor, in 2020. Her research interests include ocean remote sensing monitoring, risk assessment, and machine learning.

**YUN LI** is currently pursuing the master's degree in electronic and communication engineering with the China University of Petroleum (East China), Qingdao. Her research interests include signal detection and data processing in marine remote sensing field.

**JIANHUA WAN** received the bachelor's degree in engineering survey and the Ph.D. degree in geodesy and survey engineering from Wuhan University, Wuhan, China. He is currently a Professor and a Post-Doctoral Supervisor with the Surveying and Mapping Science and Technology, College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao. His research interests include ocean remote sensing and key technologies for smart cities.

**XINRONG LYU** received the B.Eng. degree in electronic engineering and the Ph.D. degree in pattern recognition and intelligent system from Xidian University, Xi'an, China. He is currently an Associate Professor with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao. His research interests include signal detection and data processing in ocean remote sensing field.

**PENG REN** (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees in electronic engineering from the Harbin Institute of Technology, Harbin, China, and the Ph.D. degree in computer science from the University of York, York, U.K.

He is currently a Professor with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, China. His research interests include remote sensing and machine learning.

Dr. Ren was a recipient of the K. M. Scott Prize from the University of York, in 2011, and the Eduardo Caianiello Best Student Paper Award at the 18th International Conference on Image Analysis and Processing, in 2015, as one coauthor.

**JIE SHANG** is currently an Associate Professor and the Director of the North China Sea Marine Forecasting Center, Ministry of Natural Resources (MNR), Qingdao, China. His research interests include ocean remote sensing monitoring and forecasting.