

EARTH SCIENCES

Deep-learning-based information mining from ocean remote-sensing imagery

Xiaofeng Li ^{1,2}, Bin Liu³, Gang Zheng⁴, Yibin Ren^{1,2}, Shuangshang Zhang⁵, Yingjie Liu^{1,2}, Le Gao^{1,2}, Yuhai Liu^{1,6}, Bin Zhang^{1,2} and Fan Wang^{1,2,*}

¹Key Laboratory of Ocean Circulation and Waves, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China; ²Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao 266071, China; ³College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China; ⁴State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China; ⁵College of Oceanography, Hohai University, Nanjing 210098, China and ⁶Dawning International Information Industry Co., Ltd., Qingdao 266101, China

*Corresponding author. E-mail: fwang@qdio.ac.cn

Received 11 January 2020; Revised 3 March 2020;

Accepted 6 March 2020

ABSTRACT

With the continuous development of space and sensor technologies during the last 40 years, ocean remote sensing has entered into the big-data era with typical five-V (volume, variety, value, velocity and veracity) characteristics. Ocean remote-sensing data archives reach several tens of petabytes and massive satellite data are acquired worldwide daily. To precisely, efficiently and intelligently mine the useful information submerged in such ocean remote-sensing data sets is a big challenge. Deep learning—a powerful technology recently emerging in the machine-learning field—has demonstrated its more significant superiority over traditional physical- or statistical-based algorithms for image-information extraction in many industrial-field applications and starts to draw interest in ocean remote-sensing applications. In this review paper, we first systematically reviewed two deep-learning frameworks that carry out ocean remote-sensing-image classifications and then presented eight typical applications in ocean internal-wave/eddy/oil-spill/coastal-inundation/sea-ice/green-algae/ship/coral-reef mapping from different types of ocean remote-sensing imagery to show how effective these deep-learning frameworks are. Researchers can also readily modify these existing frameworks for information mining of other kinds of remote-sensing imagery.

Keywords: ocean remote sensing, big data, artificial intelligence, image classification

INTRODUCTION

The ocean accounts for about 71% of Earth's surface. Humans had minimal ocean observations before the Seasat, the first Earth-orbiting satellite designed for remote sensing of Earth's oceans, launched in 1978 [1]. Although Seasat only operated for 105 days, the sensors on board Seasat had acquired much more data about the vast ocean than all previous sensors combined. Such high-efficiency data collection stimulated the fast development of ocean-satellite remote sensing. Since then, more and more satellites carrying microwave, visible, infrared sensors have been launched to measure various ocean physical, biological and other parameters that lead to significant improvement in our understanding of the ocean during the last 40 years [2–7].

There are two types of remote-sensing sensors: active and passive sensors. The active sensors measure sea-surface height or SSH (altimeter),

sea-surface roughness (synthetic-aperture radar or SAR), sea-surface wind (scatterometer, SAR). In contrast, the passive sensors measure sea-surface salinity, sea-surface temperature (SST) and water-leaving radiance using microwave/infrared radiometers and optical sensors. According to the report from the Committee on Earth Observation Satellites (CEOS), for each primary ocean-surface parameter, there are currently a dozen satellites on-orbit making the daily measurements (Fig. 1). Other tens of satellites have also been approved or planned over the next 20 years. The increase in satellite numbers has resulted in a rapid rise in the volume of ocean-satellite data archives that number tens of petabytes. Also, due to the improvement of spatial, temporal and spectral resolutions of various sensors, the variety of ocean-satellite data now increases. Ocean remote sensing now has the typical five-V (volume, variety, value, velocity and veracity) characteristics of big data.

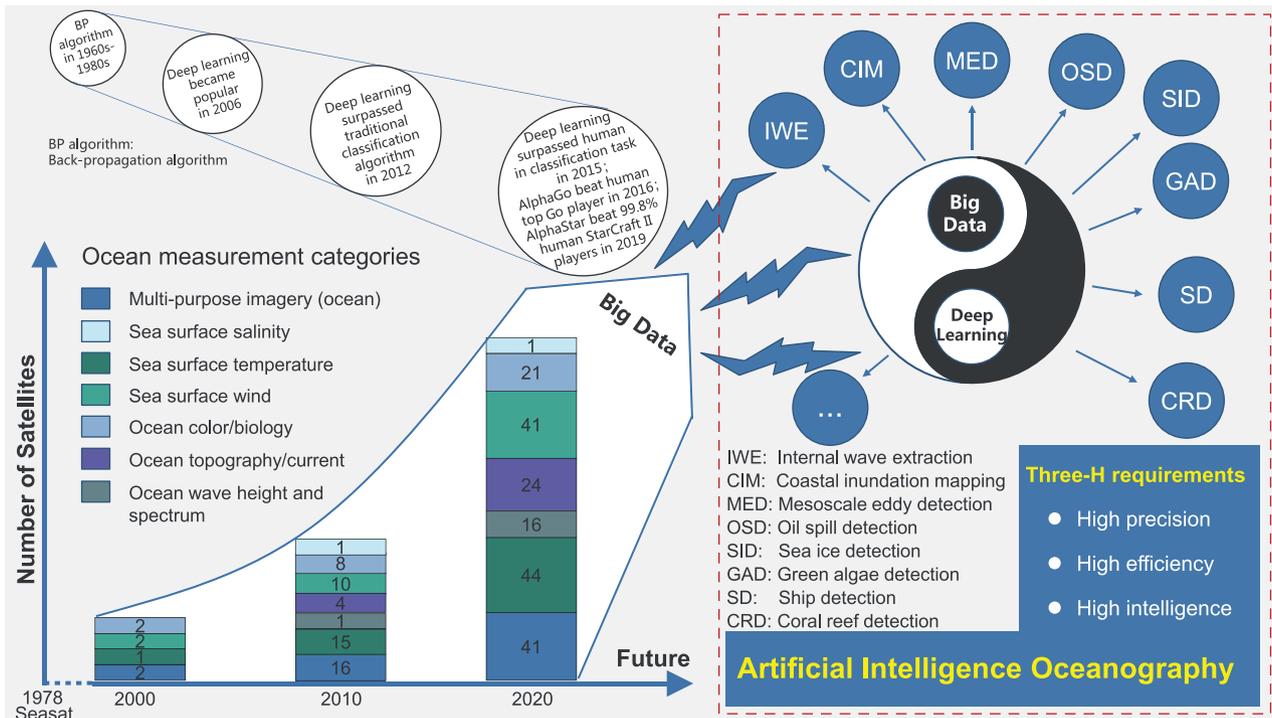


Figure 1. Ocean remote sensing has entered into the big-data era with the rapid increase in on-orbit satellites, sensors and data-archive volume. Ocean remote-sensing big data can offer abundant data fuel to data-driven deep learning, while data-driven deep learning provides a promising way to make the best of ocean remote-sensing big data. The win-win combination of them will make future ocean remote sensing more precise, efficient and intelligent. The numbers of sensors for the different ocean-measurement categories were calculated from the data of the Committee on Earth Observation Satellites (CEOS) database (<http://database.eohandbook.com/measurements/categories.aspx?measurementCategoryId=3>).

A dilemma is that big data do not always guarantee that people can get more valuable information from them because the useful information is usually sparsely hidden in massive ocean-satellite data. In the past 40 years, many efforts have been put into the tasks to develop and validate retrieval algorithms to generate a long-time series of many standard global ocean parameters [6–8]. Currently, we need efficient and even intelligent approaches to improve information-extraction capability and efficiency with emergent powerful deep-learning (DL) technology. We need to strengthen our skills in three aspects. First, some ocean phenomena like internal waves and algal blooms are locally generated and their signatures only consist of a tiny percentage of an ocean remote-sensing image. We cannot extract this type of information as we do for the standardized ocean-parameter (SST, etc.) products from the direct measurements of satellites. Second, there is much essential information hidden in these long-time series data that requires new data-driven information-mining algorithms. Besides, extracting such information from a high-rate downlink satellite data stream requires high-speed data processing. Deep-learning-based approaches can satisfy all these requirements.

Traditionally, we can categorize information extraction in ocean remote-sensing images into two types: supervised classification and object detection. Supervised classification means classifying images or pixels, usually referring to samples, according to given classes. Pixel-level supervised classification, also named supervised semantic segmentation, is more often encountered in ocean-satellite-image applications. Oil spills, sea ice, eddy and algal blooms usually have discriminable patterns in satellite images with irregular shapes [9]. Extraction of such information is an essentially supervised semantic segmentation that includes light-spectrum combination, polarization decomposition, co-occurrence matrices, spectrum analysis (e.g. wavelets) and other methods. Object detection in ocean remote-sensing imagery usually refers to detecting objects (ships, oil rigs, etc.) that are distinguished from the surrounding image backgrounds. A constant false-alarm rate (CFAR) is the most common statistical approach for ship detection in ocean remote-sensing images [10]. The methods work but may not be optimal for a specific end-to-end (data-to-information) problem, since the traditional supervised classification and object-detection approaches do not consider spatial structure features

or use the features extracted by human-designed operators.

An artificial intelligence (AI) method can help to get better and fast results. AI based on an artificial neural network (ANN) computational model that can learn the relationship among its inputs and outputs from given training samples. Biological neural networks initially inspired ANN development in the 1940s. Since then, ANN has advanced from the simple McCulloch-Pitts and Perceptron models of the 1950s–1960s and the back-propagation algorithm was developed in the 1960s–1980s (Fig. 1). The introduction of convolutional layers and pooling layers occurred in the 1980s. And the popular deep neural networks (DNNs) did not start until the early 2000s (Fig. 1). The introduction of convolutional layers primarily reduces the number of network parameters by local-linking and weight-sharing, while pooling layers reduce the size of the feature maps by down-sampling [11–13]. Please note that a DNN is different from a convolutional neural network (CNN) in concept. DNN means the neural network architecture is deep and complex, while CNN means convolutional layers are used in the neural network. When a DNN contains convolutional layers, it is also a CNN. When a CNN is deep, it is also a DNN. A deep structure with alternation of convolutional and pooling layers gives DNNs the powerful capability to efficiently extract abstract features from images. DNN training is to find the optimal structure and coefficients based on a large number of labeled samples. Once trained, a DNN can better extract the data features than the traditional approaches that use human-designed rules and then infer the information behind the data. DNNs achieved significant superiority over traditional classification approaches in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2012 and dominated the ILSVRCs in the following years. In 2015, DNNs achieved better performance (4.94% top-5 error) than human (5.1% top-5 error) on the 1000-class ImageNet 2012 classification data set [14] (Fig. 1).

U-Net is a representative DNN for semantic segmentation [15,16]. U-Net has the encoder–decoder structure and the skip connection between the encoder and the decoder, and these make U-Net have excellent performance. The DNNs for object detection can be divided into one-stage networks (e.g. single-shot multi-box detector (SSD)) and two-stage networks with a new network branch for initial screening (e.g. Faster-RCNN) [17,18]. Generally, one-stage networks have higher computational efficiency because of simpler designs. Essentially, there is no difference in information extraction between ocean remote sensing and conventional images. Therefore, powerful DNNs also have mas-

sive potential for mining information from ocean remote-sensing data and, conversely, ocean-satellite big data can provide data to fuel the DNNs.

As shown below, the information extraction of ocean remote-sensing data is undergoing the evolution from human-designed rule-based models to end-to-end learning models in the big-data era. Most of the previous ANN models applied in ocean remote sensing were based on fully connected neural networks (FNNs). The critical shortcoming of FNNs is their inefficiency in processing high-dimension data, including extracting contextual features from images. Therefore, a remedy was often made in previous FNN-based classifications by increasing an additional preprocessing step to obtain contextual features as inputs of FNN using human-designed rules. For instance, people need to extract textural features of ocean remote-sensing images in oil spills and sea-ice classification. This additional step is called feature engineering in the field of machine learning. Since this extra level uses human-designed rules, the FNN-based classification models are still not end-to-end models. The powerful capability of the feature extraction of DNNs can overcome this problem. Recent studies have shown that DNNs have achieved excellent performance in information extraction from the ocean remote-sensing data of oceans and others, although both opportunities and challenges still exist [19].

AI oceanography development is just in its infancy. And potential deep-learning applications in the oceanography field urgently need to be widely studied. Besides the classification and semantic-segmentation tasks mentioned above, deep-learning models as well as other AI models can also find their positions in observational data fusion, parameter retrieval, forecast, etc. of the ocean and atmosphere [7,20–24]. In very recent times, a deep-learning model has successfully made skillful forecasts of El Niño/Southern Oscillation, showing great potential in solving this complex scientific problem [24]. Other new advances were also achieved in applying deep-learning models to make short-term forecasts of tropical cyclone tracks and intensity [22,23].

The deep-learning frameworks can be blended. The blended models can achieve more complex functions and have improved accuracy and increased efficiency. We can also combine the deep-learning model with a physical equation-based numerical weather-prediction model. The potential combinable aspects include quality control, parameterization, post-processing, etc. Application of DL technology in weather and ocean forecasting is an up-and-coming research area, as it is possible to fulfill the advantages of both DL and numerical

modeling jointly. Recent work in the research area has been well reviewed by Boukabara *et al.* [23]. Additionally, ensemble learning for improving classification performance in the field of machine learning can also be used to combine DL models and scientists have been exploring some in remote-sensing retrievals and ocean forecasts [7,25–27].

In this review paper, we first systematically review two deep-learning frameworks that carry out ocean remote-sensing image classification and then present eight typical applications in ocean internal-wave/eddy/oil-spill/coastal-inundation/sea-ice/green-algae/ship/coral-reef mapping from different types of ocean remote-sensing imagery to show how effective these deep-learning frameworks are (Fig. 1).

DNN FRAMEWORKS OF CLASSIFICATION AND DETECTION

As mentioned above, in order to extract sparse but valuable information from a large volume of ocean remote-sensing data, we need to construct end-to-end DNN models. There are two basic tasks to mine multidimensional, or even multi-source, ocean remote-sensing data—that is, pixel-level classification and object-level detection. The applications presented in the next section can be categorized into these two tasks, although they may have different conventional terms, such as internal-wave-signature extraction, coastal-inundation mapping and mesoscale-eddy detection. In this review, for the pixel-level classification and object-level detection tasks in ocean remote-sensing data, we construct DNN models based on two classic, widely applied network frameworks—that is, U-Net [15] and SSD [17], respectively. The two frameworks are briefly introduced as follows.

Framework of U-Net

Although developed initially for the semantic segmentation of biomedical images [15], U-Net achieves successful applications in many fields. U-Net is so named because of its almost symmetric encoder–decoder network architecture that can be depicted by a U-shaped diagram (Fig. 2a). U-Net uses the skip connection to pass the intermediate feature maps extracted by the encoder to the decoder. This idea helps to reduce the information loss caused by the resolution decrease in the data stream of the encoder.

U-Net extracts the features from an input image by outputting a class confidence for each pixel. The maximum confidence of a pixel indicates the class

that the pixel is in. Then, the pixel-level classification map of the input data is generated.

As illustrated in Fig. 2a, U-Net consists of an encoder (left half in blue) and a decoder (right half in green). The encoder is used to extract image features at different resolutions. Along the data stream in the encoder, composite layers of cascaded convolutional layers alternate with max-pooling layers, and the feature-map resolution in the stream decreases after each max pooling. These composite layers can also be changed for more sophisticated ones (e.g. ResNet blocks [28]). The lowest-resolution feature maps extracted by the encoder are input into the decoder through the bottleneck at the bottom. Although those feature maps in which the grids have the largest receptive fields make the best use of the context in the input data, the more exceptional image features essential for the localization accuracy of semantic segmentation are lost due to the down-sampling caused by the max-pooling layers. To reduce the resolution loss, U-Net also passes the intermediate higher-resolution feature maps to the decoder by the skip connection (denoted by yellow modules in Fig. 2a). Contrary to the encoder, the decoder has an expansive data stream upward for resolution restoration. Its network architecture is almost the mirror of the encoder architecture, where the max-pooling operations are changed with up-sampling ones. The up-sampling operations can be realized by transposed convolutional layers or more efficient interpolation layers. After up-sampling, the decoded feature maps are concatenated with the encoder feature maps at the same resolution passed by the skip connection. Then, the concatenated feature maps are further decoded by the upper composite layer. The above procedure is repeated until the feature maps in the data stream have the same resolution as the input data. Then, these feature maps are processed by soft-max classification to yield the class confidence of each pixel.

Outlines of the object areas are delineated in raw image samples to generate pixel samples for the training of U-Net. Each pixel of the images is given a class label according to the outlines. Then, the label of each pixel is encoded into a one-hot vector that contains the ground-truth class confidence of the pixel. Such a pair of a pixel and its class confidence is a pixel sample.

The classification loss of U-Net is a soft-max loss, measuring the deviation of the output class confidences from ground-truth ones. The soft-max loss here is the sum of soft-max cross-entropy loss at each pixel. In the classic U-Net, to cause the contributions of some critical pixels to be large in the classification loss, a pixel-wise weight map is introduced in the soft-max loss, and the above sum is replaced with

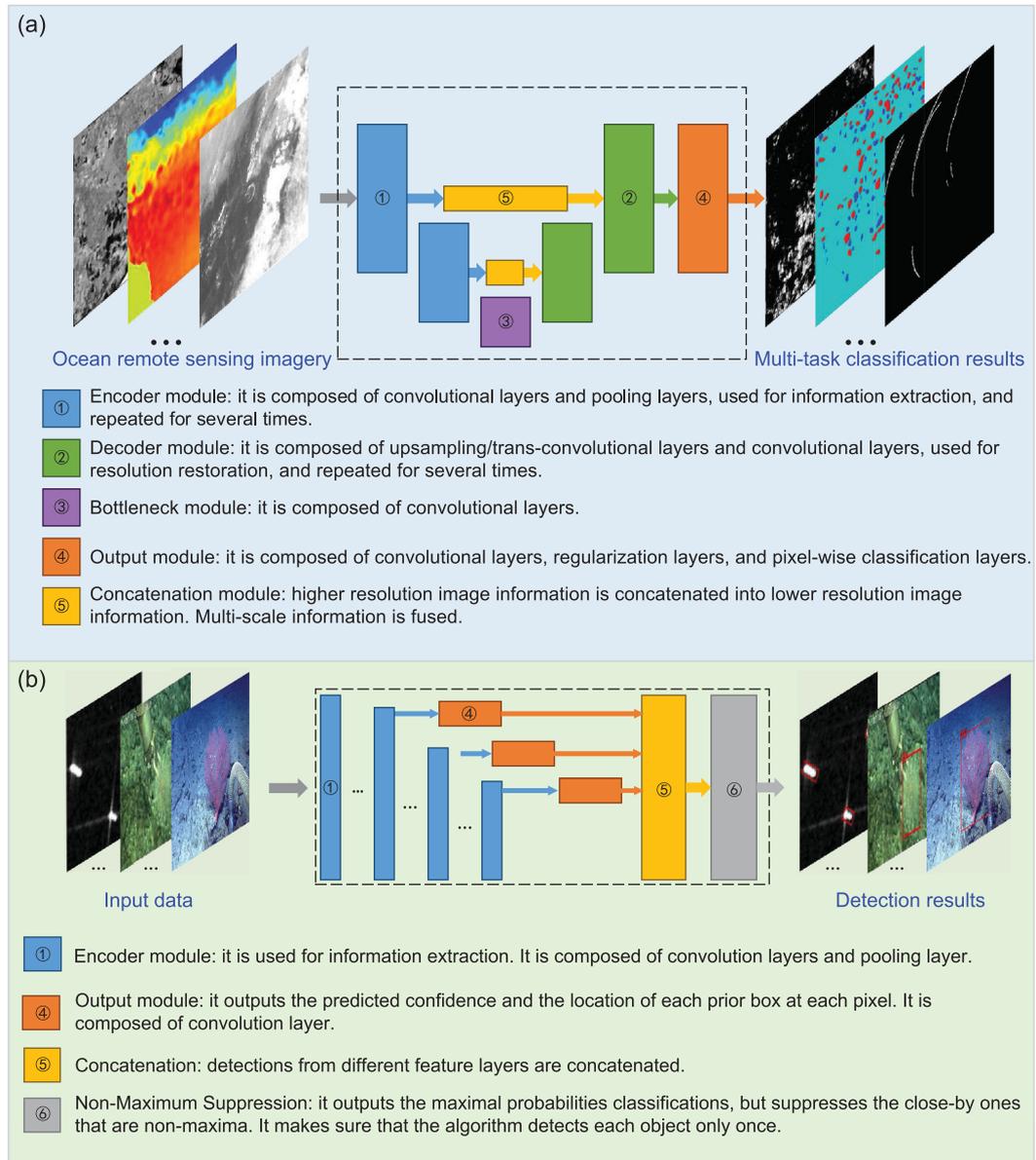


Figure 2. Moduled frameworks for pixel-level classification and object-level detection in ocean remote-sensing data. (a) The framework of U-Net. (b) The framework of single-shot detection (SSD).

the weighted sum. The above essential pixels can be the pixels that are challenging to classify or the pixels of the classes having high importance but low occurrence frequency. The weight map can be used to balance the class-occurrence frequencies. An instance of a weight map is given in [16].

Framework of SSD

SSD is a typical one-stage DNN for object detection without any fully connected layers [17]. SSD was proposed for efficiency. Tested on the VOC2007 data set, SSD run several times faster than the two-stage benchmark network Faster-RCNN [18] with a

little higher accuracy in terms of mean Average Precision [17]. Different from other networks, SSD directly makes object detection in the multi-resolution feature maps of an input image, which enables SSD to detect objects at different scales.

SSD recognizes objects in an input image and outputs the rectangular areas (boxes) occupied by the objects as well as the confidences of the objects for different classes. The maximum confidence of an output box indicates the class that the object occupying the box is in. The output boxes are represented by their encoded center locations, widths and heights. The output boxes are redundant. Therefore, after the above representation is decoded, the output boxes

are screened by non-maximum suppression to get the final boxes.

As shown in Fig. 2b, the backbone of SSD consists of two parts. The first part is to extract general features of objects from an input image, where feature-extraction modules of other networks can be adopted. In the classic SSD, VGG-16 [29] is used as the part and the fully connected layers in VGG-16 are replaced with convolutional layers. The second part is made up of several convolutional layers cascading the first part. The second part generates the multi-resolution feature maps. These multi-resolution feature maps progressively decrease in size with increasing network depth by a two-stride-step convolution. Accordingly, the receptive fields of the feature-map grids enlarge. SSD adopts a design similar to anchor boxes in Faster-RCNN [30]. Several boxes with different widths and heights are set at each grid of the multi-resolution feature maps. These boxes are called default boxes or prior boxes in SSD. The default boxes are also multi-resolution. That means that the default boxes of the feature maps at different resolutions have different scales relative to the input image. As the grid receptive fields enlarge, the box scales increase. The second part of the backbone has several network branches. The feature maps at different resolutions are further processed in the corresponding branches with small-kernel convolution, soft-max classification and bounding-box regression to get the class confidences, encoded center locations, widths and heights of the output boxes of each grid.

Ground-truth boxes and classes of objects are labeled in raw image samples to generate training samples of SSD. Then, intersection-over-union (IoU) is calculated between each ground-truth box and each default box. A boxed pair that has the maximum IoU of the ground-truth box in the pair is considered as a positive sample. The box pairs with IoU larger than a threshold (e.g. 0.5) are also considered as positive samples. Still according to the rule: a ground-truth box can match multiple default boxes in the positive samples, but a default box can only match one ground-truth box. The locations, widths and heights of the ground-truth boxes in the positive samples are encoded for the calculation of SSD loss. The rest of the default boxes are negative samples. Additionally, the strategy of hard-negative mining is also used to further balance the numbers of negative and positive samples. The strategy is that only the negative samples with the highest confidence loss are involved in the calculation of SSD loss rather than all the negative samples. The ratio between the negative and positive samples is at most 3:1.

The total SSD loss is a weighted sum of two components that are for bounding-box regression and

classification, respectively. The first component is a smooth L1 loss to measure the loss between the encoded locations, widths and heights of the output boxes and those of the ground-truth boxes in positive samples. Another component is a soft-max loss to measure the classification loss, which is the sum of the soft-max cross-entropy losses of the samples used for training. The weight in the total SSD loss is usually set at one. Some detected objects (e.g. ships) have orientations. SSD can also be designed to have orientation-detection capability by adding the orientation angles of objects in the outputs and correspondingly modifying the first SSD loss component [31,32].

APPLICATION EXAMPLES

In this section, we review the DNN-based supervised classification and object-detection applications in extracting several typical oceanic phenomena in ocean remote-sensing imagery. The applications include using geostationary satellite images for ocean internal wave information extraction; using SAR images for coastal-inundation mapping, sea-ice mapping, oil-spill mapping and ship detection; using the standard ocean remote-sensing AVISO (Archiving, Validation, and Interpretation of Satellite Oceanographic) SSH product for global ocean eddy detection; and using the MODIS (Moderate Resolution Imaging Spectroradiometer) images for Enteromorpha extraction. Using underwater-camera images, we also showed that the DNN-based model could be readily applied to extract coral-reef information from the seafloor.

Internal-wave-signature extraction

The oceanic internal wave (IW) is a ubiquitous feature of oceans. It has attracted considerable research interest because of its essential role in ocean acoustics, ocean mixing, offshore engineering and submarine navigation [33–35]. Scientists have long recognized the potential for using satellite imagery for studying IWs [36–38]. Satellite images can compensate for the *in situ* observations to study the generation, propagation, evolution and dissipation of the IWs. In the past few decades, algorithms and techniques for automated detection of IW signatures from SAR imagery using basic image-processing methods have been studied significantly [39–41]. SAR is an active sensor that measures the sea-surface roughness. It is not affected by cloud cover and can image the ocean surface at a meter to tens of meters of spatial resolution under all weather conditions, day and night. The nature of SAR coverage is

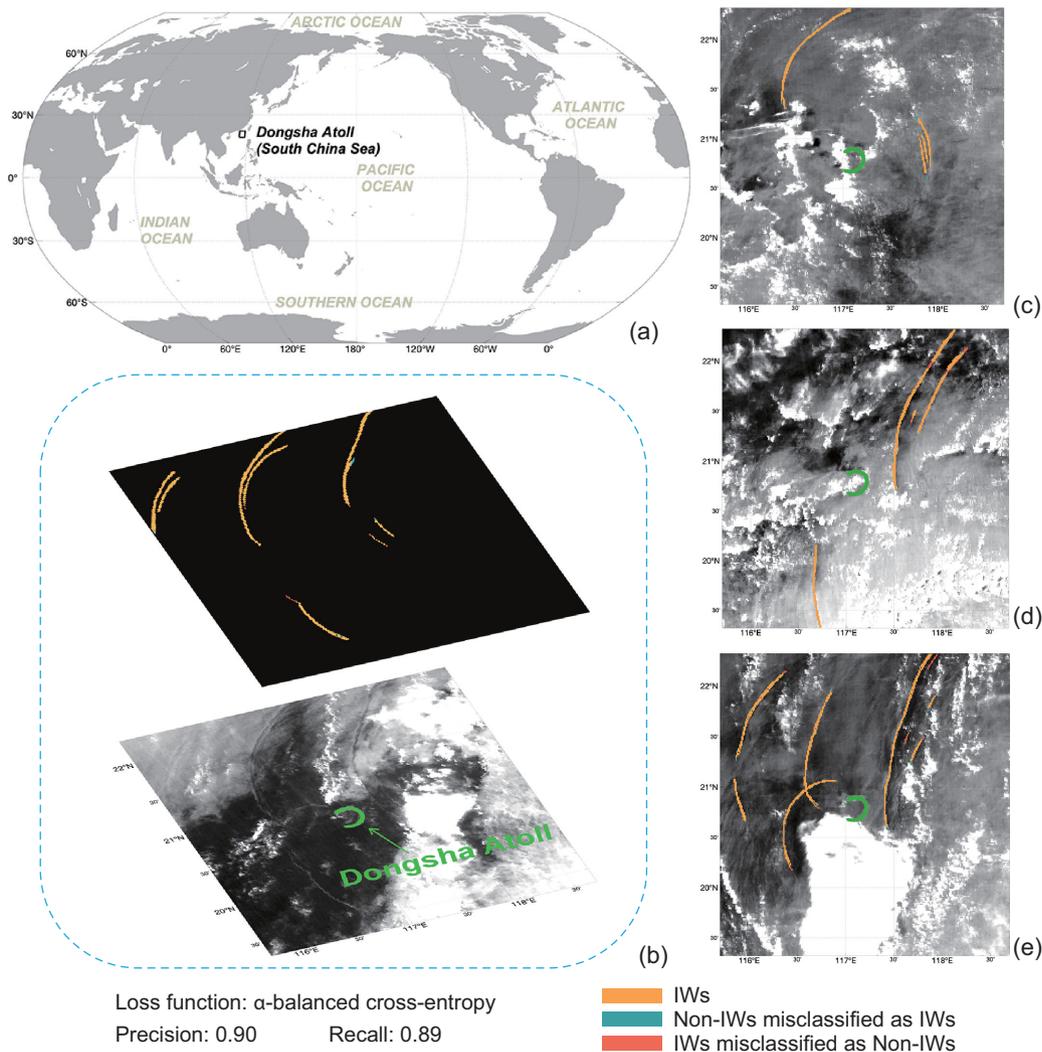


Figure 3. Four examples of the 40 testing results. (a) shows the study area. (b–e) are the input Himawari-8 images overlaid with their corresponding trained model extraction results. The four images were acquired at 05:40 on 30 May, 05:20 on 26 May, 06:00 on 21 May and 05:10 on 26 June 2018 (UTC), respectively.

low. Therefore, scientists have also tried to extract IW information from geostationary satellite images that have lower spatial (250–500 m) but higher temporal (10 minutes) resolution under suitable solar-illumination conditions. Cloud cover and solar flares make the IW signatures much weaker and more challenging to extract from geostationary satellite images than that from the SAR images [42].

In recent years, machine-learning methods have been used widely to extract robust high-level information from satellite images. In this case, we applied a modified U-Net (Fig. 2a) framework to obtain IW-signature information in Himawari-8 images under complex imaging conditions.

The Himawari-8 satellite provides visible images of Earth's surface at a spatial resolution of 0.5–1 km and a 10-minute temporal resolution [43]. It is a useful tool to monitor and investigate the IWs in

the South China Sea (SCS) [44]. We collected 160 Himawari-8 red-band images (1-km resolution) containing IW signatures around Dongsha Atoll in the northern SCS (Fig. 3) in May and June 2018.

The details of this modified U-Net framework is shown in Fig. 3. IW-signature extraction is essentially a binary pixel-wise classification problem. Typically, the loss function is a cross-entropy loss. However, in this IW-extraction case, IW only exists in very few pixels of the image, making the samples highly unbalanced and the use of cross-entropy loss invalid. To solve the class-imbalance problem, we adopted the α -balanced cross-entropy of Lin *et al.* [45] as the loss function and achieved excellent results (we set α to 0.99). To reduce the computation cost without losing generalization ability, we converted the images to the gray level and divided them into 256×256 pixel sub-images.

One hundred and twenty images, with their corresponding manually annotated ground-truth map for IW signatures (white) and surroundings (black), were randomly selected to train the U-Net framework, with the remaining 40 used as the testing images. After 200 epochs, the mean precision and recall of the testing set are 0.90 and 0.89, respectively.

Figure 3 shows the four examples from the 40 testing results. One can see that the sea-surface signatures of IWs were being not only overlaid with different types of clouds, but also strongly contaminated by solar flares as well as inhomogeneous darkness induced by other oceanic processes, all making the target signatures relatively weaker and difficult to extract. However, compared to the manually annotated ground-truth maps, the results from the U-Net model are good. Figure 3c captures a group of rarely observed reflected IWs propagating towards the east [46].

The statistical results (Fig. 3), relatively high mean precision and recall value of the 40 testing images showed that the U-Net framework is promising for the extraction of IW information in satellite images, even under complex imaging conditions.

Coastal-inundation mapping

Tropical cyclone (TC)-induced coastal inundation is a typical compound natural hazard. It is the combination of storm-surge-caused inundation and heavy-rain-induced river flooding. TC-induced coastal inundation causes a huge loss of life and property in coastal areas [47–49]. Accurate mapping of coastal inundation from remote-sensing data can not only assist in the management of performing better disaster relief, but also help researchers to better understand the inundation mechanisms and develop more accurate forecasting models. SAR is a suitable sensing means for coastal inundation because it can provide day-and-night, all-weather observation abilities and high-resolution images of the flooded areas. The traditional ways of coastal-inundation mapping from SAR images include: histogram thresholding [50], active contour segmentation [51], region growing [52], change detection [53] and statistical classification [54]. The traditional methods are based on human-crafted features or rules to mine multidimensional SAR-image data for inundation mapping. It is difficult for them to provide robust performance under several influential factors: (i) the inherent speckle noise of SAR images; (ii) SAR-system factors; (iii) meteorological influences; and (iv) environmental influential factors.

Deep convolutional neural network (DCNN) models, composed of DNN-based models with convolutional layers, can provide a promising way to

solve coastal-inundation-mapping problems. In the DCNN methods, the features for robust pattern classification for coastal-inundation mapping are mined from the multidimensional SAR data, instead of being predefined. This end-to-end, data-driven pattern-classification design is suitable for robust coastal-inundation mapping. Kang *et al.* [55] used the fully convolutional network model to verify that DCNN-based flooding detection is more accurate than the traditional methods. Rudner *et al.* [56] presented a DCNN-based method that is useful for flooded-building detection in urban areas. Liu *et al.* [57] proposed an improved DCNN method that has robust performance for coastal-inundation mapping from bi-temporal and dual-polarization SAR images.

To highlight the advantage of AI applications in coastal-inundation mapping, we constructed a DCNN framework to study this phenomenon. This framework can be generalized into studies involving multi-temporal SAR-information mining.

The framework is based on U-Net, as shown in Fig. 2a. The left part of the framework is the encoding part. It extracts abstracted features for accurate classification. It is composed of four encoding–decoding modules, as indicated by module 1/2 in Fig. 2a. Each encoding module includes two convolutional layers and one max-pooling layer. The right decoding part restores the feature-map resolution for pixel-level classification. Each decoding module includes one up-sampling layer and two convolutional layers. The input multidimensional SAR data are composed of pre-event, post-event and difference images of VH and VV polarizations (VH, vertical transmit and horizontal receive; VV, vertical transmit and vertical receive). This physics-based input information design fuses temporal and polarimetric information for more accurate coastal-inundation mapping. In the output module, as indicated by module 4 in Fig. 2a, a SpatialDropout2D layer is added before the classification to regularize the model for better generalization abilities [58]. Since inundation mapping is a binary classification problem, the binary cross-entropy is used for the loss function [45]. The detailed model design can be found in [57].

We used the 10-m-resolution Sentinel-1 SAR data to perform the experiments. We applied radiometric calibration, 7×7 boxcar filtering and geocoding to the SAR images. The ground-truth labels were extracted from Copernicus Emergency Management Service Rapid Mapping products [59] and human delineation with the help of Google Earth and OpenStreetMap.

The training/testing samples were generated from 2017 Harvey-induced coastal inundation in Houston, Texas, USA and from 2019 Lekima-induced coastal inundation in Zhejiang,

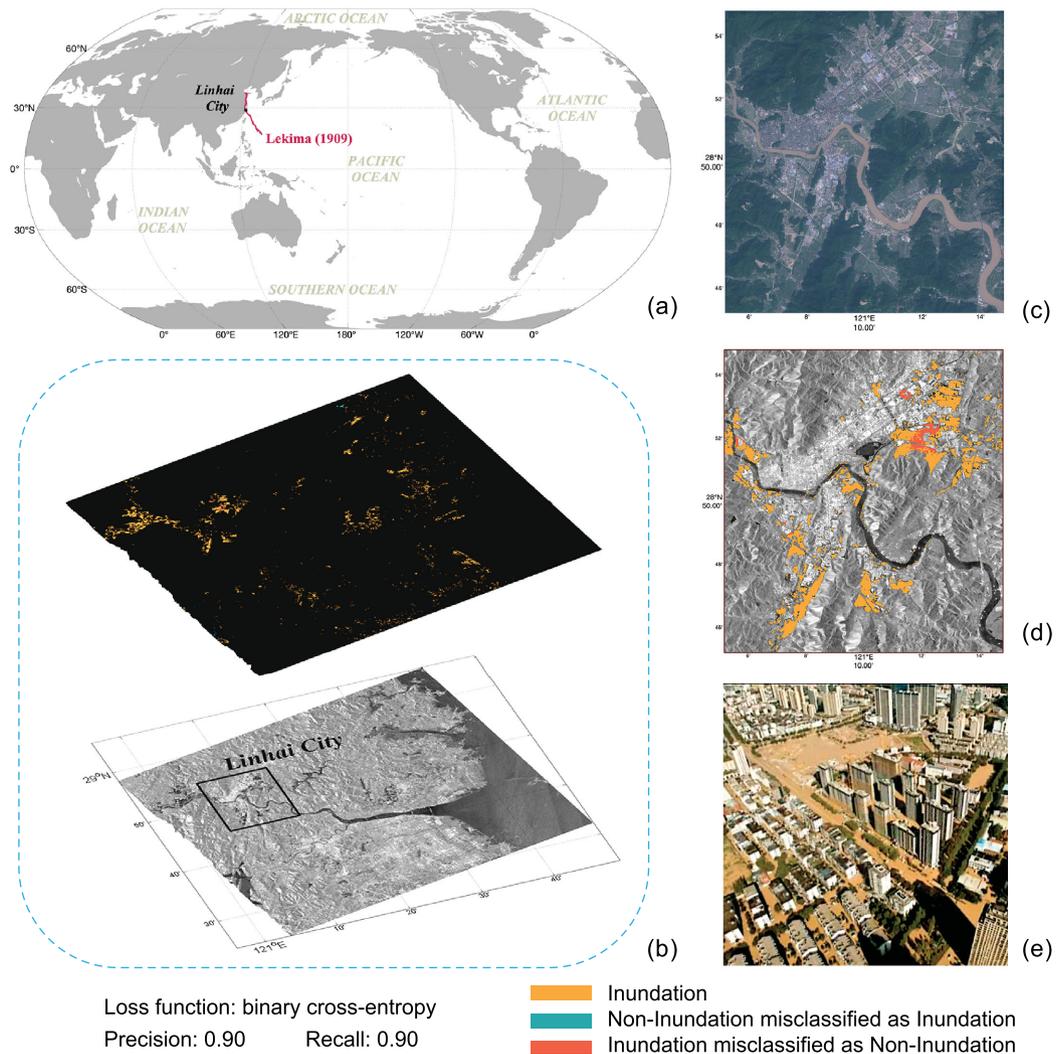


Figure 4. Deep-learning-based mapping result of Lekima-caused coastal inundation in Zhejiang, China. (a) The path of Lekima 2019. (b) The mapping result (upper image) and the pre-event SAR image (lower image) of the scene around Linhai city, Zhejiang. (c) The Sentinel-2 optical image of Linhai city. (d) The mapping result superimposed onto the SAR image in Linhai. (e) Picture taken after the passage of Lekima shows severe flooding in Linhai.

China, respectively. The 2019 Lekima path and the mapping result are shown in Fig. 4a and b. The orange areas are correctly extracted inundation. The red areas are missed detections and the cyan areas are false alarms. The precision and recall of the result are both around 0.90. The result shows the severe inundation in Linhai city, Zhejiang, as illustrated in Fig. 4d. The Sentinel-2 optical image of Linhai is shown in Fig. 4c. The picture taken after the passage of Lekima confirms that Linhai was severely flooded, as shown in Fig. 4e.

The AI technology, particularly the DCNN methods, can mine multidimensional SAR data for accurate and robust coastal-inundation mapping. In the future, the model can be extended to work under multiple image-source conditions for more practical applications.

Global mesoscale-eddy detection

Mesoscale eddies are circular currents of water bodies that widely exist in the global oceans. They play a significant role in the transport of momentum, mass, heat, nutrients, salt and other seawater chemical elements across the ocean basins. They also impact global ocean circulation, large-scale water distribution and biological activities effectively [60–63].

Automatic eddy-identification algorithms include the physical-parameter-based method [64–66], the flow-direction-based method [67] and the SSH-based method [61]. All these algorithms lack the computational efficiency of contour iterations or have complex calculation processes.

In the introduction part, we described that the DNN-based framework can solve many practical

problems such as pattern recognition and computer vision with high efficiency. It is natural to propose using the DNN framework to detect mesoscale eddies that have prominent patterns in the global SSH maps. In the literature, Lguensat *et al.* [68] developed ‘EddyNet’ that was based on the encoder–decoder network U-Net (Fig. 2a) to detect oceanic eddies in the southwest Atlantic. Franz *et al.* [69] also use the U-Net architecture to identify and track oceanic eddies in Australia and the East Australia current. Du *et al.* [70] developed ‘DeepEddy’ based on PCANet (Principal Component Analysis Network) and spatial pyramid pooling to detect oceanic eddies based on SAR images. Xu *et al.* [71] applied the pyramid scene parsing network ‘PSPNet’ to catch eddies in the North Pacific Subtropical Countercurrent region. These regional studies proved that the DNN performed well in detecting mesoscale eddies in territorial seas. The DNN performance on global mesoscale-eddy detection remained unverified.

In this section, we applied a generalized DNN framework to detect global mesoscale eddies. The framework is based on the U-Net architecture (Fig. 2a) consisting of ResNet blocks. Each ResNet block is composed of a 3×3 convolutional layer, followed by a batch normalization (BN) and a rectified linear unit (ReLU) activation, then a 3×3 convolutional layer and a BN layer. The output is added to the input to be activated by a ReLU layer. For the encoder path, each block is followed by 2×2 max pooling and dropouts. For the decoder path, transposed convolutions are used to restore the original resolution. The dice loss function, which is widely used in segmentation problems, is the cost function. The $0.25^\circ \times 0.25^\circ$ spatial resolution and daily SSH product during 2000–11 was generated by Ssalto/Duacs and distributed by AVISO. Mesoscale eddies that identified by the SSH-based method [61] during the same period were used as the ground-truth data set. Pixels were labeled as ‘1’, ‘-1’ and ‘0’ inside anticyclonic eddies (AEs), cyclonic eddies (CEs) and background regions.

Figure 5a shows the mesoscale eddies identified by the DNN method on 1 January 2019. There are a total of 3314 (2963 ground-truth) AEs and 3407 (3056 ground-truth) CEs in the global ocean. Compared to the SSH-based method, the accuracy of the DNN-based eddy-detection method is 93.79% and the mean IoU is 88.86%. Figure 5b clearly shows that the DNN-based framework identified many more small-scale eddies. Besides, it takes <1 minute for the DNN-based method costs to identify eddies in the global ocean, while the SSH-based method takes >16 hours [72].

The satellite-tracked drifter is used to validate the eddy-detection results of the DNN-based method. The drifters have their drogues centered at 15-m depth to measure the surface currents and make either a cycloidal or a looping trajectory when trapped in an eddy. As shown in Fig. 5c, a drifter (ID: 43677) was trapped in a CE in the eastern North Pacific on 1 May 2011. After 20 days, the drifter captured in the CE moved as a counterclockwise loop (Fig. 5d). Another two counterclockwise loops of drifter trajectory can be seen in the CE in Fig. 5e and f, after 25 and 30 days, respectively. Such a result is consistent with the concept that CEs rotate counterclockwise in the Northern Hemisphere.

In conclusion, the DNN-based eddy-detection method can not only identify many more small-scale eddies, but also significantly improve the computational efficiency. Further development of the DNN-based framework includes adding other types of ocean remote-sensing images, i.e. SST, chlorophyll concentration, etc., to a multi-parameter-based DNN framework.

Oil-spill detection

Oil spill is typical marine pollution. Oil floating on the sea and beaching on the shore could seriously affect surrounding marine fisheries, aquaculture, wildlife, ecosystems, maritime tourism and transportation, among others. For example, the Deepwater Horizon (DWH) oil spill is an example of severe marine-pollution disasters. It happened in the Gulf of Mexico on 20 April 2010 and an estimated $7.0 \times 10^5 \text{ m}^3$ of oil was released into the sea before the well was capped on 15 July 2010 [73]. Accurate detection of oil spills from remote-sensing data can help disaster-relief managers to perform targeted responses and it can also assist scientists in forecasting the movement and dissipation processes of oil spills. SAR is a suitable sensor for oil-spill detection because the oil dampens the sea-surface capillary waves so that they appear dark in SAR intensity images [74–76]. Besides, oil slicks also modulate the surface-scattering matrix received by advanced polarization SAR [77–83]. As a result, oil slicks also have significant signatures in the full-polarization SAR images.

In the big-ocean-data era, AI technology has the potential to mine information from a high volume of polarization SAR images acquired under different meteorological conditions and system parameters. It is promising to develop a robust feature-detection algorithm using such technology. For example, Chen *et al.* [84] used the DNN framework to optimize the polarimetric SAR feature sets to reduce the feature

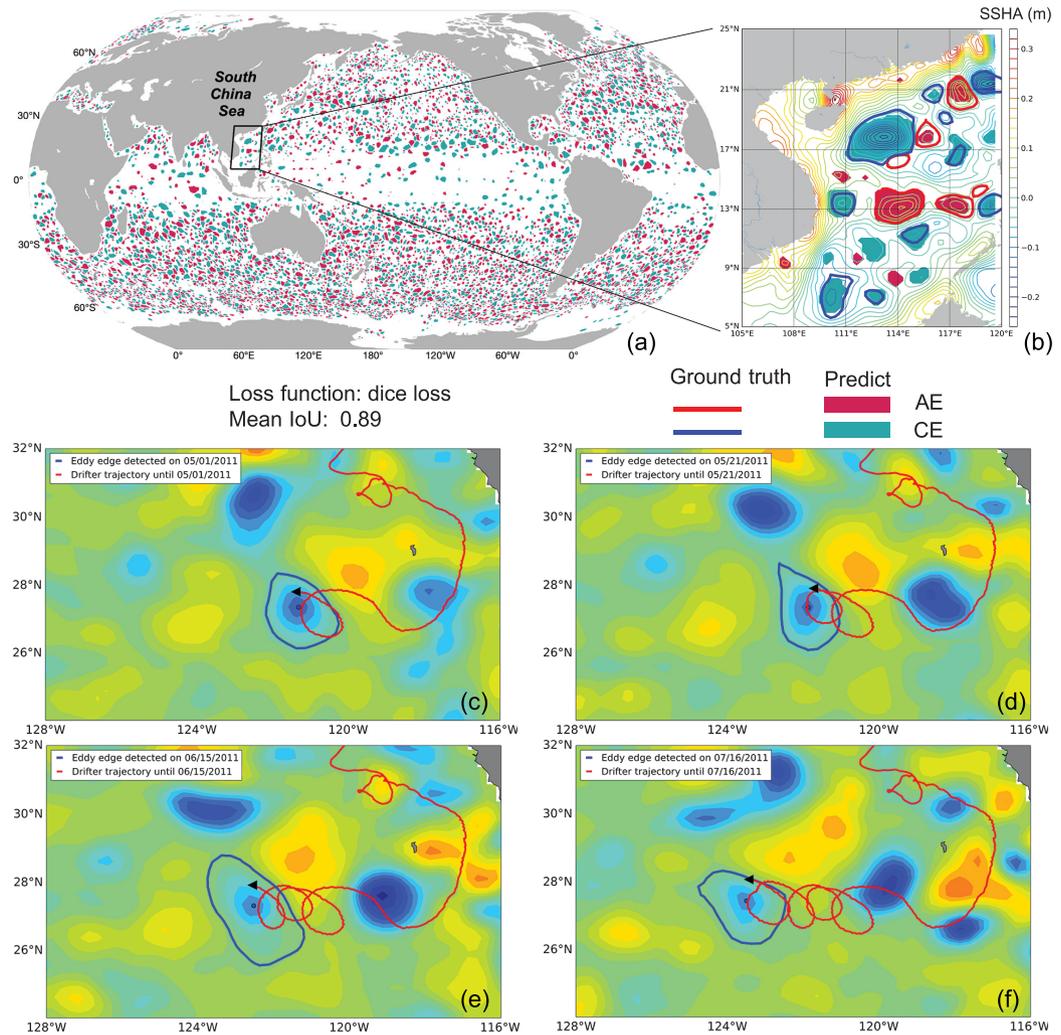


Figure 5. (a) Mesoscale eddies detected by the AI-based method in the global ocean on 1 January 2019. (b) Mesoscale eddies detected by the SSH-based method and AI-based method in the SCS on 1 January 2019. A drifter is captured by a CE that was detected by the AI-based method in the eastern North Pacific and rotated with the CE on (c) 1 May 2011, (d) 21 May 2011, (e) 15 June 2011 and (f) 17 July 2011 (the color denotes SSHA — sea surface height anomaly).

dimension for oil-spill detection and classification. Guo *et al.* [85] proposed a DNN-based polarization SAR-image-discrimination method for oil slicks and lookalikes. Guo *et al.* [86] used the DNN-based semantic-segmentation model to detect oil spills by using backscattering energy information.

To demonstrate that the AI technology has great potential for robust oil-spill detection and characterization under various meteorological and SAR-acquisition conditions, we constructed a generalized AI framework to study oil detection in polarization SAR data.

The framework is based on U-Net, as shown in Fig. 2a. The left encoding part extracts abstracted features with four encoding modules, each including two convolutional layers and one max-pooling layer. The right decoding part restores the feature-map res-

olution with four decoding modules, each including one up-sampling layer and two convolutional layers. The input configuration is to use the diagonal components of the polarimetric coherence matrix [87], T11, T22, T33, with and without the incidence angle. In the output module, a Spatial-Dropout2D layer [57] is added before the classification to regularize the model for better generalization. Since oil-spill detection is a binary classification problem, we selected the binary cross-entropy as a loss function [45].

We applied this DNN-based model to a set of L-band Uninhabited Aerial Vehicle SAR (UAVSAR) images taken during the DWH oil-spill event by the National Aeronautics and Space Administration (NASA) [88]. The UAVSAR is a full-polarization SAR with fine resolution (7 m),

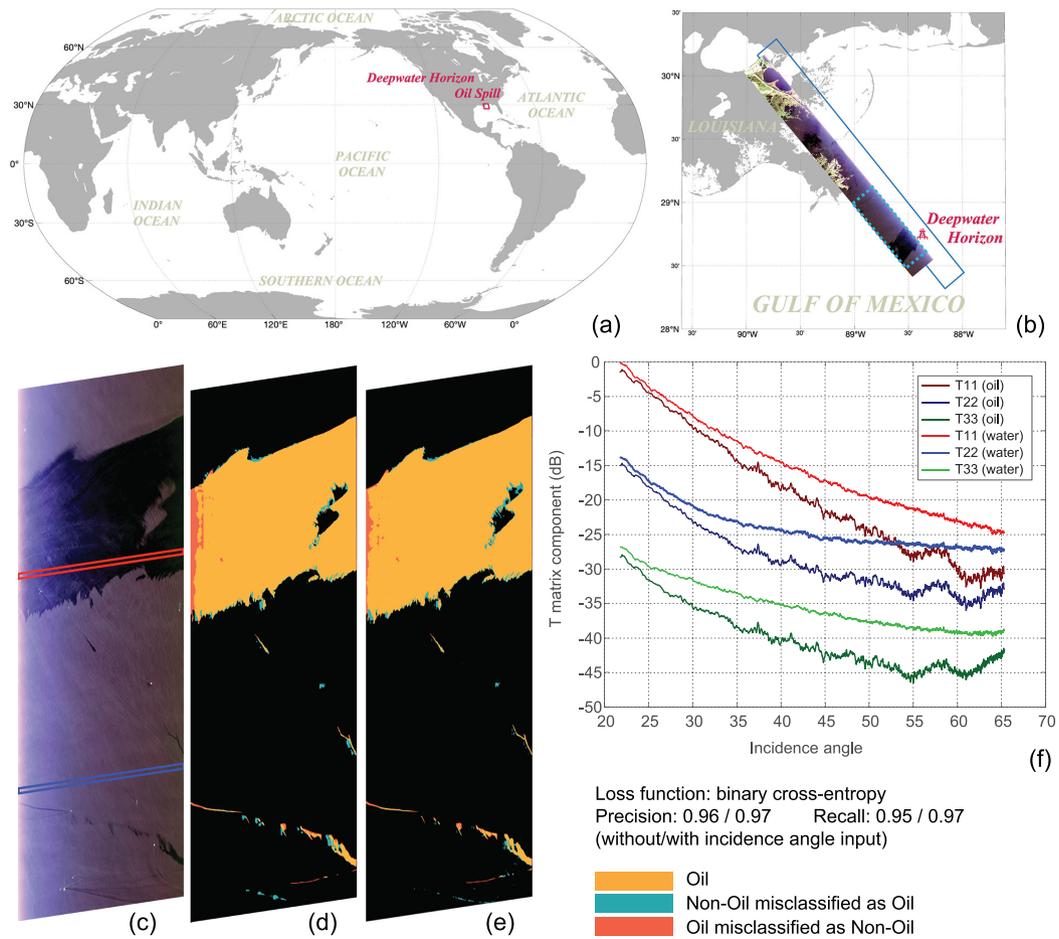


Figure 6. Deep-learning-based mapping result of the Deepwater Horizon (DWH) oil spill in the Gulf of Mexico. (a) The DWH oil spill in the world map. (b) The polarimetric SAR images used in the training and testing. The polarimetric SAR image in the dashed blue rectangle was used for testing and performance evaluation. (c) The PauliRGB pseudo-color image of the testing area. (d) The deep-learning-based result without the incidence angle. (e) The deep-learning-based result with the incidence angle. (f) The relationship of the polarimetric coherence matrix components and the incidence angle of the oil and water areas in (e).

stable calibration and low noise floor. Radiometric calibration and 7×7 boxcar filtering were applied to the data. To show the incidence angle's influence, geocoding was not used. According to the detailed analysis with *in situ* observations of [89], we manually extracted ground-truth labels. The training/testing samples were from different UAVSAR images with flight ID 14010/32010.

Figure 6 shows the testing image with PauliRGB pseudo-color [87]. From left to right, the incidence angle increases from 22° to 65° . We can observe that the incidence angle has an impact on the oil–water discrimination capability. An oil area (in the red rectangle) and a water area (in the blue box) were selected to show the trends of the T11, T22 and T33 components with incidence angle, as illustrated in Fig. 6c. We can see that the incidence angle influences the oil–water discrimination capability. Below

30° , it is challenging to discriminate oil from water. Moreover, according to [89], the backscattered information is influenced by the noise floor, especially above 60° . The detection results without and with incidence angle are shown in Fig. 6d and e, respectively. The correctly detected areas are in orange, the missed detections are in red and the false alarms are in cyan. The recall and precision of the result without using the incidence angle are 0.95 and 0.96, respectively. The recall and accuracy of the result with the incidence angle are both improved to 0.97. With the incidence angle in training and testing, the AI technology can mine reliable features for robust pattern classification, even with a minimal incidence angle.

AI technology, particularly the DCNN methods, can mine multi-polarization SAR data for accurate oil-spill detection. It shows the potential for robust detection under various influential factors. In the

future, a more extensive range of factors should be considered and analysed for practical use.

Sea-ice detection

Sea ice is a significant threat to marine navigation and transportation safety. The change of sea-ice distribution reflects the interaction of the atmosphere–cryosphere–hydrosphere and global climate change. Sea-ice detection and monitoring draw attention widely. SAR, which is independent of sun illumination and cloud conditions, plays a vital role in sea-ice monitoring [90]. A series of studies have been devoted to the SAR sea-ice-detection problem. The critical challenge is to develop a robust model that captures domain-specific expert knowledge for discriminating between ice and water using SAR backscatter characteristics. To achieve this goal, different types of sea-ice-detection models based on backscatter thresholding [91], regression techniques [92], expert systems [93], Bayesian techniques [94], gray-level co-occurrence matrix and the support vector machine (SVM) hybrid method [95], among others [96], are proposed.

Recently, with the rapid progress of AI technology, researchers have employed DNN to extract features automatically to improve the accuracy and efficiency of sea-ice classification. Xu and Scott [97] introduced an earlier CNN-based model AlexNet and transferred learning to classify sea ice and open water. Gao *et al.* [98] integrated transfer learning and dense CNN blocks to form a transferred multi-level fusion network (MLFN). The MLFN outperformed the PCAKM [99], the NBRELM [100] and the GaborPCANet [101] in classifying sea ice and open water. Similar DL-based studies were carried out by other researchers [102,103]. More and more researchers are trying to construct DL-based models to achieve end-to-end sea-ice detection with higher accuracy and stability.

To highlight the advantage of AI applications in sea-ice classification, we constructed a generalized AI framework based on U-Net (Fig. 2a). The input was a 256×256 SAR image. We stacked four encoder modules to extract features level by level. Each encoder was composed of two convolutional layers with ReLU units followed by a max-pooling layer. One bottleneck module that was composed of two stacking convolutional layers with ReLU units was added onto the last encoder module. Four decoder modules were stacked upon the bottleneck module. Each decoder module was composed of one up-sampling layer and two stacking convolutional layers with ReLU units. The concatenation module was applied to fuse the encoder module and the decoder

module at the same level. The output module consisted of one CNN layer with one activation layer, which outputted the predicted value of each pixel. We transformed the detection procedure as a binary classification problem: sea ice or open water. Thus, we applied sigmoid as the activation function. If the predicted value is >0.5 , the pixel belongs to sea ice. Otherwise, the pixel is open water. The loss function is binary cross-entropy.

To show the effectiveness of this AI model, we acquired six Sentinel-1 SAR images in the Bering Strait between February and April 2019. The 10-m-resolution VV channel of the ground range detected SAR product was selected as the experiment data set. We scaled all SAR pixel values between 0 (water) and 1 (sea ice). Each SAR image was divided into small chips with a size of 256×256 pixels.

There were 1340 chips in the training set. The batch size was 16. The initial learning rate was 0.0001. We split 20% samples from the training set as the validation set. The early stopping strategy was adopted to avoid overfitting. Finally, the model ran 86 training epochs. The precision and the recall of the testing set were 0.95 and 0.91, respectively. As shown in Fig. 7b–e, most of the sea ice, including small chunks and sinuous ice edges, could be successfully detected. However, the rough sea surfaces resulted in some misclassifications, which need to be further addressed.

The proposed U-Net-based model is capable of detecting sea ice from SAR images at the pixel level. The detection framework is an end-to-end model without manual feature engineering and expert knowledge. The detection results (Fig. 7) show that details such as the boundary between sea ice and open water can be successfully detected. In the future, employing DNN-based models to detect or estimate more parameters of sea ice, such as the type, the thickness and the intensity, etc., should be new challenges.

Green-algae detection

Enteromorpha prolifera (EP), as a kind of large-sized green algae, is fulminant and drift in the East and Yellow Seas of China in spring and summer seasons. Sporadic EP first occurred along the coast of Jiangsu province at about $35^{\circ}54'$ in Fig. 8a, then drifted northward driven by the wind and current. During the drifting process, when meeting the appropriate water-quality conditions, large-scale proliferation and aggregation occurred. Since 2008, a high concentration of EP has beached along the coast every year, causing a so-called green-tide disaster affecting ship traffic, the environment, coastal ecosystems, public health and the tourism

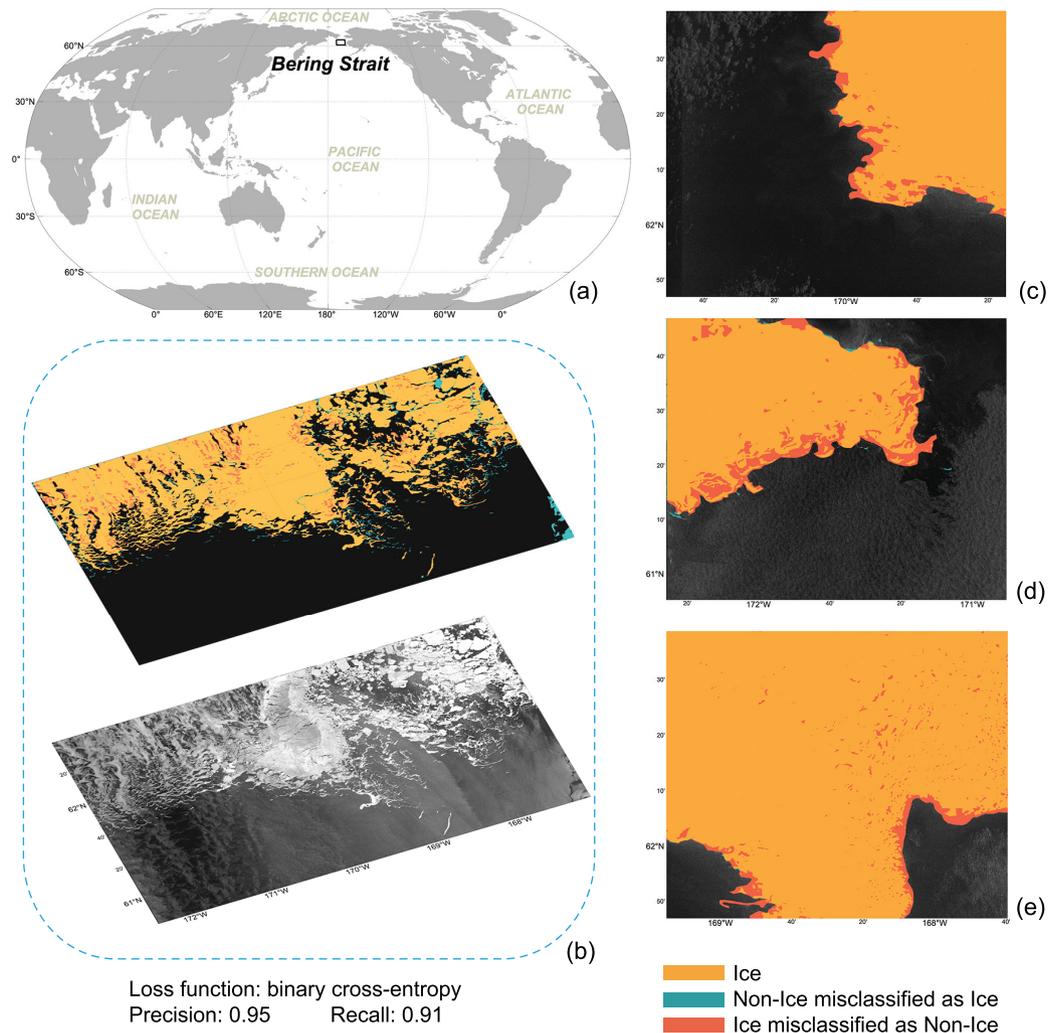


Figure 7. Testing results of the proposed sea-ice-detection framework. (a) The overall location of the study area. (b) Detection results of the first testing SAR image. (c–e) Detection results of the second testing SAR image.

industry, among others [104,105]. Mapping and tracking EP in near real time facilitates treatment processing.

EP has distinguished features in sea-surface-reflectance images acquired by the MODIS sensors on board NASA Aqua and Terra satellites. The aggregation of seaweed and its decomposition alter the water-surface-reflectance values [106]. For EP and other types of floating green-algae detection in ocean remote-sensing images, multiband ratio methods, e.g. NDVI (normalized difference vegetation index [107]) and FAI (floating algae index [108]) were developed. In general, these methods have reasonable visual interpretation and low error rates.

As we pointed out in the introduction part, DNN has significant superiority over the physical-based algorithms for image classification. For example, Arellano-Verdejo *et al.* [109] proposed a DNN framework model named ‘ERISNet’ to classify the

presence/absence of pelagic sargassum along the coastline of Mexico. The model is based on a 1D CNN and achieves a probability of 90.08% in the classification at the pixel level. In this section, we customized an EPNet model based on the U-Net framework (Fig. 2a) for EP classification in MODIS imagery. The significant difference between ‘ERISNet’ and ‘EPNet’ is that we used 2D convolution in the DNN model. The intermediate architecture consists of symmetrical ascending and descending structures, which include five encoder and decoder modules, respectively, and the binary cross-entropy is the loss function.

We manually extracted EP labels from the MODIS true-color images (bands: 1/4/3). To construct a labeled data set, we collected different types of tags (banded, lumpy and dotted types) under different environmental conditions. We also followed the common practice of expanding the

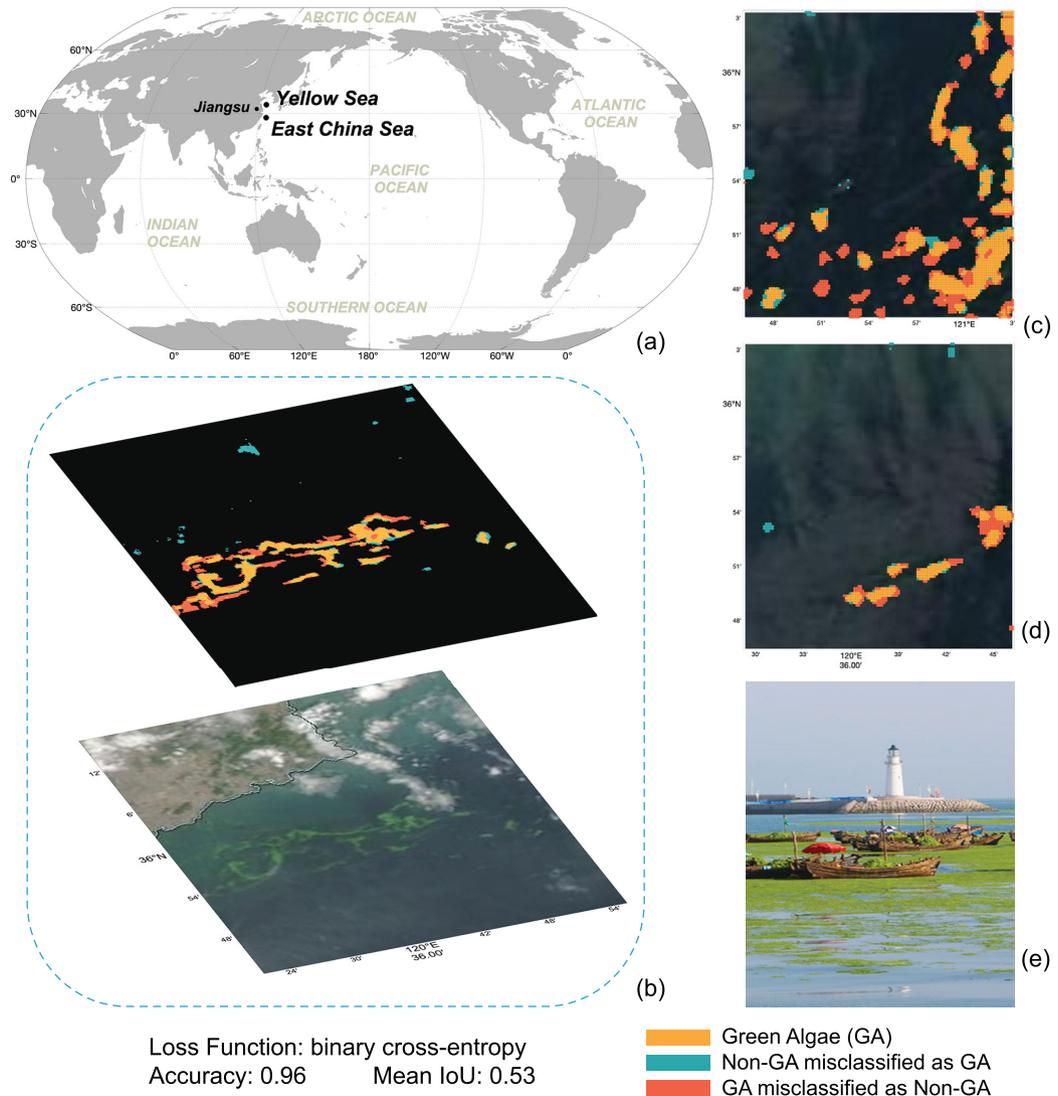


Figure 8. *Enteromorpha prolifera* (a type of green algae) detection. (a) Study area. (b) MODIS images and its corresponding *Enteromorpha prolifera* classification result. (c) and (d) The fallout ratio or the omission ratio in some pixels due to the cloud contamination of the optical images. (e) Picture taken after *Enteromorpha prolifera* bloom.

sample size by rotation of the sample images by 90° , 180° and 270° . For MODIS images acquired between 2008 and 2019, we obtained 1680 pairs of MODIS EP slices and their corresponding labels (128×128 pixels). A randomly selected 1460 and 220 pairs were used as training and testing sets, respectively.

Figure 8a shows the EP blooms in the Yellow Sea of China. EPNet achieves an overall classification accuracy of 0.96 with a mean IoU of 0.53 (Fig. 8b). However, there are some fallout ratios or omission ratios in some pixels in Fig. 8c and d. The misclassification is due to the cloud contamination of the optical images. Nevertheless, our analysis shows that the DNN framework can be readily implemented to identify EP.

Ship detection

Ship detection plays a significant role in marine surveillance. SAR has been widely used in marine-ship detection because it is capable of monitoring ocean targets under all weather conditions, day and night [110–112]. For decades, a series of studies have been devoted to detecting ships and other targets in SAR images. The algorithms can be divided into conventional methods and AI-based methods. A typical conventional method is threshold-based methods that focus on finding bright pixels operating accurate clutter statistical modeling. Algorithms built on the theory of CFAR filtering [10] and generalized likelihood ratio testing (GLRT) [113] are representations. The main drawback of the conventional methods is that they need prior professional

knowledge to manually design features, which has been a common challenge faced by most fields in the era of big data [11].

The cutting-edge AI framework, deep learning, can extract features automatically, which has been a great achievement in computer vision. Faster region-based convolutional network (Faster-RCNN) [18] is a complete end-to-end CNN target-detection model. Researchers have introduced Faster-RCNN to detect ships in SAR images [114]. Other studies tried to introduce rotatable bounding boxes into detection models to represent ships more accurately. Liu et al. [115] proposed a detector using rotatable bounding boxes (DRBox), which optimized the traditional SSD [17] by rotating the prior box. DRBox outperformed traditional SSD and Faster-RCNN in detecting densely arranged ships. Other rotated detectors such as rotation dense feature pyramid networks (R-DFPN) [31] and DRBox-v2 [116] were proposed successively. Notably, DRBox-v2 improved DRBox by integrating new DL tricks such as a feature pyramid network (FPN) [117] and focal loss (FL) [45], which outperformed R-DFPN and DRBox in detecting ships and airplanes [116]. Compared with conventional methods, the DNN-based ship-detection models significantly simplify feature engineering and achieve end-to-end detection with higher accuracy and stability.

To highlight the advantage of AI applications in ship detection, we constructed a generalized AI framework based on SSD (Fig. 2b) to fulfill the task.

The input was a 300×300 pixels SAR image. We stacked five encoder modules to extract features. The output module performed convolution on feature maps and generated the location and the confidence of the detected boxes. Different from DRBox-v2, the detected boxes were generated based on the last four feature maps (DRBox-v2 is the last three). The new added shallow feature map helps to detect small targets. FPN was adopted to fuse features at different levels and the rotation box was adopted. The loss function (L) consists of two parts: the confidence loss (L_{conf}) and the location loss (L_{loc}):

$$L = \frac{1}{N_1 + N_2} L_{conf} + \frac{1}{N_2} L_{loc}, \quad (1)$$

where N_1 is the number of negative samples and N_2 is the number of positive samples. L_{conf} is the cross-entropy between the output and the ground truth, which can be defined as follows:

$$L_{conf} = - \sum_{i \in Pos} \log(c_i) - \sum_{i \in Neg} \log(1 - c_j), \quad (2)$$

where c_i is the confidence of the i_{th} positive samples and c_j is the confidence of the j_{th} negative samples; Pos and Neg are positive and negative sets, respectively. Hard-negative mining (HNM) [17] and FL are employed to overcome the problem of the imbalance between the positive and negative samples. For location loss, L_{loc} measures the location difference between the predicted rotation bounding box and the matched ground truth, which can be calculated as follows:

$$L_{loc} = \sum_{i \in Pos} \sum_{j \in Grd} \sum_{k \in \{x,y,l,w,\theta\}} I_{ij} smooth_{L1} \times (p_i^k - g_j^k), \quad (3)$$

where k is the location vector consisting of center coordinates (x, y), length (l), width (w) and angle (θ) of the box. p_i is the prediction location of the i_{th} positive box and g_j corresponds to the j_{th} ground truth. $smooth_{L1}$ means $L1$ norm. Grd is the ground-truth sets. I_{ij} means if the i_{th} prior box matches the j_{th} ground truth: $I_{ij} = 1$ when matching, otherwise $I_{ij} = 0$.

The OpenSARShip [118], a data set dedicated to Sentinel-1 ship interpretation, was employed as the sample data set. We labeled the ship by a rotation bounding box with a MATLAB tool shared in DRBox-v2.

The training and testing sets included 1600 and 338 ship chips, respectively. The batch size was eight. After 8200 training epochs, the loss value was <0.01 and the model stopped. Finally, 297 ships were successfully detected by the proposed model. The mAP (mean Average Precision) and mean IoU were 0.86 and 0.68, respectively. The original DRBox-v2 detected 267 ships and its mAP and mean IoU were 0.75 and 0.66, respectively. As shown in Fig. 9, some ships that were not successfully detected by the original DRBox-v2 model were successfully detected by the optimized one.

The proposed SSD-based framework is capable of detecting ships from Sentinel-1 images. By adding a shallow feature map, the detection accuracy was improved significantly. In the future, constructing end-to-end AI-based models to identify the type and geometric parameters of ships from SAR images should draw more attention.

Coral-reef detection

Marine species are a vital part of the ocean and play an essential role in the trophic chain of the ecosystems. Detection and identification of marine species are one of the crucial ways to explore marine biodiversity. Traditional methods of biometric

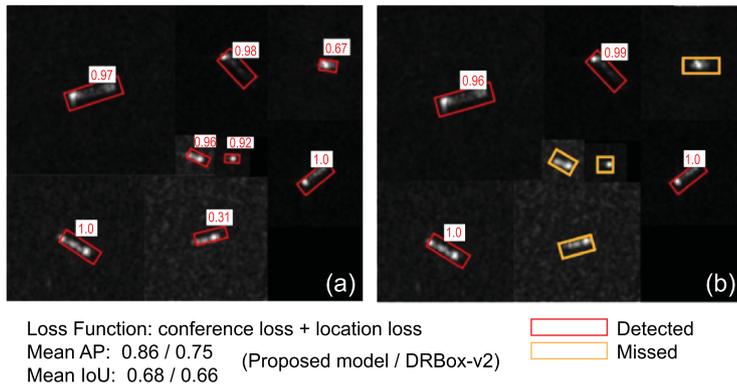


Figure 9. Testing results of the proposed ship-detection framework. (a) Detection results of the proposed optimized DRbox-v2 model. (b) Detection results of the original DRbox-v2 model. The red rectangle represents correct detection and the orange rectangle represents missed detection. The number in each white box represents the detection confidence.

identification are based on morphology and molecular genetics, or even the start-of-the-art DNA (deoxyribonucleic acid) sequencing under electron microscopy in the laboratory. Although these methods are accurate, we can not carry out such experiments in the actual marine environment. Another issue is that *ex situ* detection often causes organisms to become inactive or die.

To solve these issues, we can apply AI-based methods to detect marine species on the fly. As we have shown in previous sections, DNN has a significant advantage in satellite-image classification. The same technology can also be applied to underwater-camera-image classification. Recently, Villon *et al.* [119] used the CNN framework to detect and classify fish and they showed the CNN outperformed the traditional SVM classifier. Xu *et al.* [120] presented a comprehensive review of the computer-vision techniques for marine-species recognition from the perspectives of both classification and detection. They further compared the machine-learning techniques with deep-learning techniques and discussed the complementary issues between these two approaches. Using a new genetic-programming approach, Marini *et al.* [121] achieved high reliability when tracking fish variations at different time scales but failed to classify fish while monitoring. Saqib *et al.* [122] adopted an end-to-end Faster-RCNN for surveillance and population estimation of marine animals. Faster-RCNN significantly improves the mAP, but the detection speed is prolonged. Pedersen *et al.* [123] used the YOLOv3 (You only look once) [124] framework and the Brackish Dataset to detect big fish, small fish, crab, jellyfish, shrimp and starfish.

To highlight the advantage of AI applications, we constructed a generalized AI framework based on

SSD (Fig. 2b) for coral-reef detection in underwater-camera images. SSD is a one-stage classifier. We can train an SSD model by simultaneously optimizing classification loss and localization loss. Compared with the two-stage classifier, SSD is much faster while still ensuring classification accuracy [17]. As a result, SSD makes real-time underwater-species detection and classification possible. Our SSD framework is based on VGG16 [29], which is pre-trained on the ILSVRC CLS-LOC data set [125]. We converted fc6 (the sixth fully connected layer) and fc7 to convolutional layers, subsample parameters from fc6 and fc7 and after that removed all the dropout layers and the fc8 layer using SSD Weighted Loss Function [17]. We adjusted the outcome model using stochastic gradient descent (SGD) with an initial learning rate of 0.0004, 0.9 momenta, 0.0005 weight decay and a batch size of 32.

The experimental marine organisms included *Chrysogorgia ramificans* [126], *Chrysogorgia binata* [126], *Paragorgia rubra*, [127] and *Metallogorgia macrospina*, which were collected by RV (research vessel) *KEXUE*. We split videos of these four species into frames and annotated pictures manually. In the data preprocessing, we referred to the Australian benthic data set [128]. We randomly divided the samples into training, validation and test according to a certain proportion. To make the SSD-based model robust, we selected training-set images containing different types of species collected under different underwater conditions and shooting angles. Data preprocessing included two parts: data-format conversion and image-size standardization.

We applied the SSD-based model to 59 test images. Among all the four different coral species, Fig. 10 shows that the SSD-based model achieved 0.96 mAP with an average IoU value of 0.79. This result is remarkable and the SSD-based model can be used in coral-reef classification in real time. To further detect small-sized coral reef, we need to increase the sample size and species categories.

CONCLUSION AND FUTURE PERSPECTIVES

Ocean remote sensing has entered the big-data era with typical five-V characteristics. In the age, ideal data-mining technology should be able to extract sparse but valuable information from enormous ocean remote-sensing data volumes precisely, efficiently and with very little human involvement. The technology should also be smart and robust enough to cope with various problems that ocean remote-sensing big data contain. The above requirements can be summarized into three Hs

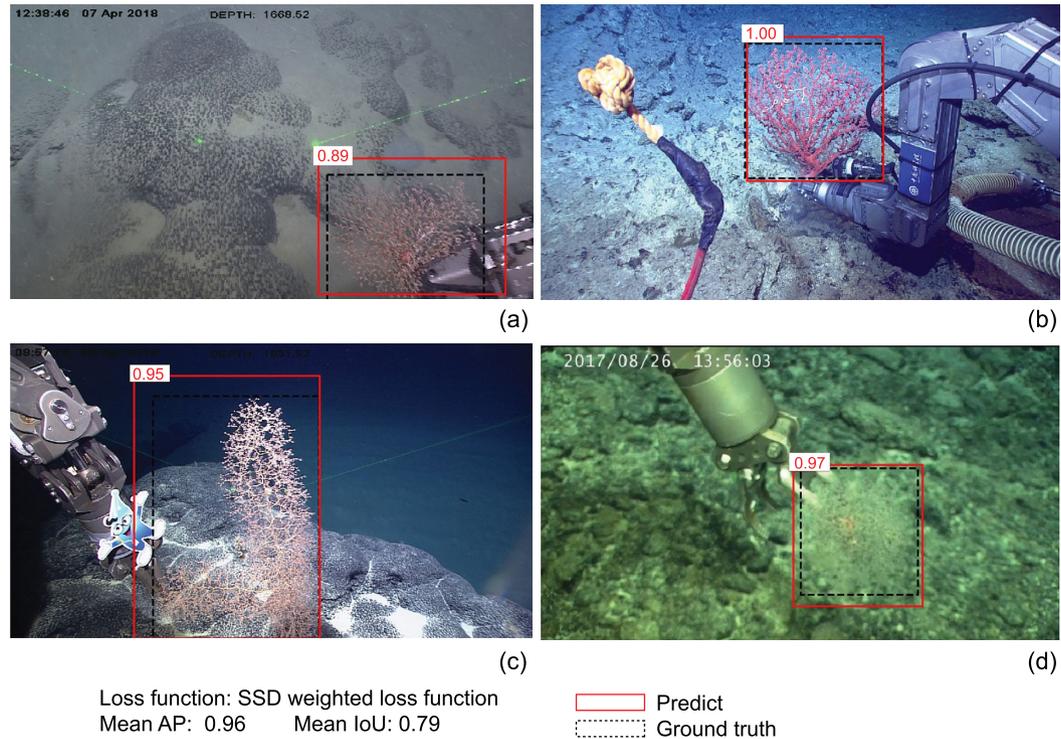


Figure 10. (a) *Chrysogorgia binata* [126], (b) *Paragorgia rubra* [127], (c) *Chrysogorgia ramificans* [126], (d) *Metallogorgia macrospina*. The number in each white box represents the detection confidence.

(high precision, high efficiency and high intelligence). Emergent deep-learning technology satisfies the three-H requirements and provides a promising way for such information extraction.

Pixel-level classification and object-level detection are two fundamental tasks in information extraction. As a result, we introduced two representative DL frameworks (U-Net and SSD) to demonstrate the powerful capability of DL technology to fulfill these tasks.

Although DL is a potent tool and demonstrates its advantages for information mining from ocean remote-sensing imagery, we still need to consider some key issues when we look ahead.

First, the DL-based technology is data-hungry. Notably, the DL-based technology needs enormous amounts of highly accurate labels. Currently, objects of interest are usually manually labeled and the label accuracy is subject to human experience and errors. As a result, the DL model trained with those labels inevitably introduces errors into the output results (e.g. statistics of shapes, sizes and areas of objects). For DL technology to reveal oceanic reality, we should provide labels from reliable *in situ* measurements, which requires world-level collaborations. Standard data sets with joint efforts of the entire community should boost the DL-based ocean remote-sensing imagery-information mining. If big

data are the door to AI ocean remote sensing, then DL technology is the key to the door. For some studies, we still rely on expert knowledge to provide the ground truth. It is essential to combine the knowledge of different expert groups to eliminate human bias. One possible solution is to develop unsupervised DL methods that avoid the limitations of human knowledge.

Second, most DL models for ocean remote-sensing imagery-information mining come from the computer-vision community. These models are developed initially to extract spatial and temporal patterns to solve vision problems. These models should and could be guided and specially tailored to serve the purpose of ocean-science applications. To tackle a specific problem, the combination of the knowledge of big-data scientists and particular domain scientists would help to reveal the real world more effectively than ever before.

Third, for DL-based ocean remote-sensing imagery-information mining, the trained DL models are often sensitive to sensors, as shown in this study. If we train these different models for different sensors, it is computationally expensive and labor-costly. We need to study practical ways of transferring models from one sensor to another and improve the model's generalization capability under different sensing conditions.

In this paper, we reviewed eight typical DL framework applications in ocean internal-wave/eddy/oil-spill/coastal-inundation/sea-ice/green-algae/ship/coral-reef mapping from different types of ocean remote-sensing imagery. We described the general deep-learning model set up for data mining in ocean remote sensing and showed that the U-Net and SSD models achieved superior performances in these topics.

ACKNOWLEDGEMENTS

The authors would like to thank the European Space Agency for providing the Sentinel-1 data, the Sentinel Application Platform (SNAP) software and the PolSARpro software; thank the Copernicus Emergency Management Service for providing the Rapid Mapping products; thank the European Space Agency (ESA) for providing the Sentinel-2 data (downloaded from Google Earth Engine); thank NASA's Jet Propulsion Laboratory for providing the UAVSAR data; and thank the Zhejiang Online News for providing Fig. 4e (<https://www.zjol.com.cn>). The Himawari-8 data are distributed by the Japan Meteorological Agency (JMA) and acquired by the HimawariCast service at Hohai University. The AVISO SSH data are produced and distributed by Aviso+ (<https://www.aviso.altimetry.fr/>), as part of the Ssalto ground-processing segment; the satellite-tracked drifter data were provided by the GDP Drifter Data Assembly Center (DAC) at NOAA's Atlantic Oceanographic and Meteorological Laboratory (www.aoml.noaa.gov/envids/gld/dirkrig/parttrk.spatial.temporal.php); the sea-ice data set was provided by ESA; the OpenSAR-SHIP was downloaded from the 'OpenSAR Platform' (<http://opensar.sjtu.edu.cn>). Ground-truth labels were annotated using the LabelMe tool (<http://labelme.csail.mit.edu/Release3.0/>). Sentinel-1 SAR radiometric calibration and terrain correction were made using SNAP 3.0 (<https://step.esa.int/main/snap-3-0-is-now-available/>).

FUNDING

The work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA19060101 and XDA19090103), the Senior User Project of RV KEXUE, managed by the Center for Ocean Mega-Science, Chinese Academy of Sciences (KEXUE2019GZ04), the Key R & D Project of Shandong Province (2019JZZY010102), the Key Deployment Project of Center for Ocean Mega-Science, CAS (COMS2019R02), the CAS Program (Y9KY04101L) and the China Postdoctoral Science Foundation (2019M651474 and 2019M662452).

Conflict of interest statement. None declared.

REFERENCES

1. Stewart RH. Seasat: results of the mission. *Bull Amer Meteorol Soc* 1988; **69**: 1441–7.
2. Li X, Pichel WG and He M *et al.* Observation of hurricane-generated ocean swell refraction at the Gulf Stream north wall with the RADARSAT-1 synthetic aperture radar. *IEEE Trans Geosci Remote Sensing* 2002; **40**: 2131–42.
3. Li X, Li C and Xu Q *et al.* Sea surface manifestation of along-tidal-channel underwater ridges imaged by SAR. *IEEE Trans Geosci Remote Sensing* 2009; **47**: 2467–77.
4. Li X, Zheng W and Yang X *et al.* Sea fetch observed by synthetic aperture radar. *IEEE Trans Geosci Remote Sensing* 2016; **55**: 272–9.
5. Zheng G, Yang J and Liu AK *et al.* Comparison of typhoon centers from SAR and IR images and those from best track data sets. *IEEE Trans Geosci Remote Sensing* 2015; **54**: 1000–12.
6. Zheng G, Li X and Zhou L *et al.* Development of a gray-level co-occurrence matrix-based texture orientation estimation method and its application in sea surface wind direction retrieval from SAR imagery. *IEEE Trans Geosci Remote Sensing* 2018; **56**: 5244–60.
7. Zheng G, Yang J and Li X *et al.* Using artificial neural network ensembles with Cropping resampling technique to retrieve sea surface temperature from HY-2A scanning microwave radiometer data. *IEEE Trans Geosci Remote Sensing* 2018; **57**: 985–1000.
8. Yang X, Li X and Pichel WG *et al.* Comparison of ocean surface winds from ENVISAT ASAR, MetOp ASCAT scatterometer, buoy measurements, and NOGAPS model. *IEEE Trans Geosci Remote Sensing* 2011; **49**: 4743–50.
9. Garcia-Pineda O, MacDonald IR and Li X *et al.* Oil spill mapping and measurement in the Gulf of Mexico with textural classifier neural network algorithm (TCNNA). *IEEE J Sel Top Appl Earth Observ Remote Sens* 2013; **6**: 2517–25.
10. Wackerman CC, Friedman KS and Pichel WG *et al.* Automatic detection of ships in RADARSAT-1 SAR imagery. *Can J Remote Sens* 2001; **27**: 568–77.
11. LeCun Y, Bengio Y and Hinton G. Deep learning. *Nature* 2015; **521**: 436–44.
12. Jordan MI and Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science* 2015; **349**: 255–60.
13. Reichstein M, Camps-Valls G and Stevens B *et al.* Deep learning and process understanding for data-driven Earth system science. *Nature* 2019; **566**: 195–204.
14. He K, Zhang X and Ren S *et al.* Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: *Proceedings of the IEEE International Conference on Computer Vision 2015*. Santiago: IEEE, 1026–8.
15. Ronneberger O, Fischer P and Brox T. U-Net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention 2015*. Munich: Springer, 234–7.
16. Falk T, Mai D and Bensch R *et al.* U-Net: deep learning for cell counting, detection, and morphometry. *Nat Methods* 2019; **16**: 67–70.
17. Liu W, Anguelov D and Erhan D *et al.* SSD: single shot multibox detector. In: *European Conference on Computer Vision 2016*. Amsterdam: Springer, 21–37.
18. Ren S, He K and Girshick R *et al.* Faster R-CNN: towards real-time object detection with region proposal networks. In: *Advances in Neural Information Processing Systems 2015*. Montréal: Curran Associates, 91–8.

19. Li J, Huang X and Gong J. Deep neural network for remote sensing image interpretation: status and perspectives. *Natl Sci Rev* 2019; **6**: 1082–6.
20. Scher S and Messori G. Weather and climate forecasting with neural networks: using general circulation models (GCMs) with different complexity as a study ground. *Geosci Model Dev* 2019; **12**: 2797–809.
21. Weyn JA, Durran DR and Caruana R. Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data. *J Adv Model Earth Syst* 2019; **11**: 2680–93.
22. Cloud KA, Reich BJ and Rozoff CM et al. A feed forward neural network based on model output statistics for short-term hurricane intensity prediction. *Weather Forecast* 2019; **34**: 985–97.
23. Boukabara SA, Krasnopolsky V and Stewart JQ et al. Leveraging modern artificial intelligence for remote sensing and NWP: benefits and challenges. *Bull Amer Meteorol Soc* 2019; **100**: ES473–91.
24. Ham YG, Kim JH and Luo JJ. Deep learning for multi-year ENSO forecasts. *Nature* 2019; **573**: 568–72.
25. Foroozand H, Radić V and Weijs SV. Application of entropy ensemble filter in neural network forecasts of tropical Pacific sea surface temperatures. *Entropy* 2018; **20**: 207.
26. Kussul N, Lavreniuk M and Skakun S et al. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci Remote Sens Lett* 2017; **14**: 778–82.
27. Kim J, Kim K and Cho J et al. Satellite-based prediction of Arctic sea ice concentration using a deep neural network with multi-model ensemble. *Remote Sens* 2019; **11**: 19.
28. He K, Zhang X and Ren S et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016*. Las Vegas: IEEE, 770–8.
29. Simonyan K and Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556.
30. Nie GH, Zhang P and Niu X et al. Ship detection using transfer learned single shot multi box detector. In: *ITM Web of Conferences*. San Diego: EDP Sciences, 01006.
31. Yang X, Sun H and Fu K et al. Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks. *Remote Sens* 2018; **10**: 132.
32. Liu Z, Hu J and Weng L et al. Rotated region based CNN for ship detection. In: *2017 IEEE International Conference on Image Processing (ICIP)*. Beijing: IEEE, 900–4.
33. Haury LR, Briscoe MG and Orr MH. Tidally generated internal wave packets in Massachusetts Bay. *Nature* 1979; **278**: 312–7.
34. Zheng Q, Susanto RD and Ho CR et al. Statistical and dynamical analyses of generation mechanisms of solitary internal waves in the northern South China Sea. *J Geophys Res* 2007; **112**: C03021.
35. Li X, Zhao Z and Pichel WG. Internal solitary waves in the northwestern South China Sea inferred from satellite images. *Geophys Res Lett* 2008; **35**: L13605.
36. Zheng Q, Yuan Y and Klemas V et al. Theoretical expression for an ocean internal soliton synthetic aperture radar image and determination of the soliton characteristic half width. *J Geophys Res* 2001; **106**: 31415–23.
37. Li X, Jackson CR and Pichel WG. Internal solitary wave refraction at Dongsha Atoll, South China Sea. *Geophys Res Lett* 2013; **40**: 3128–32.
38. Dong D, Yang X and Li X et al. SAR observation of eddy-induced mode-2 internal solitary waves in the South China Sea. *IEEE Trans Geosci Remote Sensing* 2016; **54**: 6674–86.
39. Rodenas JA and Garello R. Wavelet analysis in SAR ocean image profiles for internal wave detection and wavelength estimation. *IEEE Trans Geosci Remote Sensing* 1997; **35**: 933–45.
40. Rodenas JA and Garello R. Internal wave detection and location in SAR images using wavelet transform. *IEEE Trans Geosci Remote Sensing* 1998; **36**: 1494–507.
41. Simonin D. The automated detection and recognition of internal waves. *Int J Remote Sens* 2009; **30**: 4581–98.
42. Lindsey DT, Nam S and Miller SD. Tracking oceanic nonlinear internal waves in the Indonesian seas from geostationary orbit. *Remote Sens Environ* 2018; **208**: 202–9.
43. Bessho K, Date K and Hayashi M et al. An Introduction to Himawari-8/9—Japan's new-generation geostationary meteorological satellites. *J Meteorol Soc Jpn* 2016; **94**: 151–83.
44. Gao Q, Dong D and Yang X et al. Himawari-8 geostationary satellite observation of the internal solitary waves in the South China Sea. *Int Arch Photogr Remote Sens & Spatial Inf Sci* 2018; **42**: 363–70.
45. Lin TY, Goyal P and Girshick R et al. Focal loss for dense object detection. In: *Proceedings of the IEEE International Conference on Computer Vision 2017*. Honolulu: IEEE, 2980–8.
46. Bai X, Li X and Lamb KG et al. Internal solitary wave reflection near Dongsha Atoll, the South China Sea. *J Geophys Res Oceans* 2017; **122**: 7978–91.
47. Woodruff JD, Irish JL and Camargo SJ. Coastal flooding by tropical cyclones and sea-level rise. *Nature* 2013; **504**: 44–52.
48. Patricola CM and Wehner MF. Anthropogenic influences on major tropical cyclone events. *Nature* 2018; **563**: 339–46.
49. Zhang W, Villarini G and Vecchi GA et al. Urbanization exacerbated the rainfall and flooding caused by hurricane Harvey in Houston. *Nature* 2018; **563**: 384–8.
50. Chini M, Hostache R and Giustarini L et al. A hierarchical split-based approach for parametric thresholding of SAR images: flood inundation as a test case. *IEEE Trans Geosci Remote Sensing* 2017; **55**: 6975–88.
51. Horritt M. A statistical active contour model for SAR image segmentation. *Image Vision Comput* 1999; **17**: 213–24.
52. Matgen P, Hostache R and Schumann G et al. Towards an automated SAR-based flood monitoring system: lessons learned from two case studies. *Phys Chem Earth* 2011; **36**: 241–52.
53. Bazi Y, Bruzzone L and Melgani F. An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images. *IEEE Trans Geosci Remote Sensing* 2005; **43**: 874–87.
54. Giustarini L, Hostache R and Kavetski D et al. Probabilistic flood mapping using synthetic aperture radar data. *IEEE Trans Geosci Remote Sensing* 2016; **54**: 6958–69.
55. Kang W, Xiang Y and Wang F et al. Flood detection in Gaofen-3 SAR images via fully convolutional networks. *Sensors* 2018; **18**: 2915.
56. Rudner TG, Rußwurm M and Fil J et al. Multi3Net: segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery. In: *Proceedings of the AAAI Conference on Artificial Intelligence 2018*. Honolulu: AAAI Press, 702–7.
57. Liu B, Li X and Zheng G. Coastal inundation mapping from bi-temporal and dual-polarization SAR imagery based on deep convolutional neural networks. *J Geophys Res Oceans* 2019; **124**: 9101–3.
58. Tompson J, Goroshin R and Jain A et al. Efficient object localization using convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015*. Boston, MA: IEEE, 648–8.

59. List of EMS Rapid Mapping Activations. Copernicus Emergency Management Service website, 2020. <https://emergency.copernicus.eu/mapping/list-of-activations-rapid> (23 April 2020, date last accessed).
60. Chelton DB, Gaube P and Schlax MG *et al.* The influence of nonlinear mesoscale eddies on near-surface oceanic chlorophyll. *Science* 2011; **334**: 328–32.
61. Chelton DB, Schlax MG and Samelson RM. Global observations of nonlinear mesoscale eddies. *Prog Oceanogr* 2011; **91**: 167–216.
62. Dong C, McWilliams JC and Liu Y *et al.* Global heat and salt transports by eddy movement. *Nat Commun* 2014; **5**: 3294.
63. Zhang Z, Wang W and Qiu B. Oceanic mass transport by mesoscale eddies. *Science* 2014; **345**: 322–4.
64. Chaigneau A. Mesoscale eddies off Peru in altimeter records: identification algorithms and eddy spatio-temporal patterns. *Prog Oceanogr* 2008; **79**: 106–19.
65. Chelton DB, Schlax MG and Samelson RM *et al.* Global observations of large oceanic eddies. *Geophys Res Lett* 2007; **34**: L15606.
66. Doglioli A, Blanke B and Speich S *et al.* Tracking coherent structures in a regional ocean model with wavelet analysis: application to Cape Basin eddies. *J Geophys Res* 2007; **112**: C05043.
67. NencioLi F, Dong C and Dickey T *et al.* A vector geometry-based eddy detection algorithm and its application to a high-resolution numerical model product and high-frequency radar surface velocities in the Southern California Bight. *J Atmos Oceanic Technol* 2010; **27**: 564–79.
68. Lguensat R, Sun M and Fablet R *et al.* EddyNet: a deep neural network for pixel-wise classification of oceanic eddies. In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2018*. Valencia: IEEE, 1764–3.
69. Franz K, Roscher R and Milioto A *et al.* Ocean eddy identification and tracking using neural networks. In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2018*. Valencia: IEEE, 6887–3.
70. Du Y, Song W and He Q *et al.* Deep learning with multi-scale feature fusion in remote sensing for automatic oceanic eddy detection. *Inf Fusion* 2019; **49**: 89–99.
71. Xu G, Cheng C and Yang W *et al.* Oceanic eddy identification using an AI scheme. *Remote Sens* 2019; **11**: 1349.
72. Liu Y, Chen G and Sun M *et al.* A parallel SLA-based algorithm for global mesoscale eddy identification. *J Atmos Oceanic Technol* 2016; **33**: 2743–54.
73. Crone TJ and Tolstoy M. Magnitude of the 2010 Gulf of Mexico oil leak. *Science* 2010; **330**: 634.
74. Hu C, Li X and Pichel WG *et al.* Detection of natural oil slicks in the NW Gulf of Mexico using MODIS imagery. *Geophys Res Lett* 2009; **36**: L01604.
75. Cheng Y, Li X and Xu Q *et al.* SAR observation and model tracking of an oil spill event in coastal waters. *Mar Pollut Bull* 2011; **62**: 350–63.
76. Li X, Li C and Yang Z *et al.* SAR imaging of ocean surface oil seep trajectories induced by near inertial oscillation. *Remote Sens Environ* 2013; **130**: 182–7.
77. Zhang B, Li X and Perrie W *et al.* Compact polarimetric synthetic aperture radar for marine oil platform and slick detection. *IEEE Trans Geosci Remote Sensing* 2016; **55**: 1407–23.
78. Buono A, Nunziata F and Migliaccio M *et al.* Polarimetric analysis of compact-polarimetry SAR architectures for sea oil slick observation. *IEEE Trans Geosci Remote Sensing* 2016; **54**: 5862–74.
79. Migliaccio M, Nunziata F and Brown CE *et al.* Polarimetric synthetic aperture radar utilized to track oil spills. *Eos Trans AGU* 2012; **93**: 161–2.
80. Liu P, Li X and Qu JJ *et al.* Oil spill detection with fully polarimetric UAVSAR data. *Mar Pollut Bull* 2011; **62**: 2611–8.
81. Zhang B, Perrie W and Li X *et al.* Mapping sea surface oil slicks using RADARSAT-2 quad-polarization SAR image. *Geophys Res Lett* 2011; **38**: L10602.
82. Liu P, Zhao C and Li X *et al.* Identification of ocean oil spills in SAR imagery based on fuzzy logic algorithm. *Int J Remote Sens* 2010; **31**: 4819–33.
83. Migliaccio M, Nunziata F and Montuori A *et al.* A multifrequency polarimetric SAR processing chain to observe oil fields in the Gulf of Mexico. *IEEE Trans Geosci Remote Sensing* 2011; **49**: 4729–37.
84. Chen G, Li Y and Sun G *et al.* Application of deep networks to oil spill detection using polarimetric Synthetic Aperture Radar Images. *Appl Sci* 2017; **7**: 968.
85. Guo H, Wu D and An J. Discrimination of oil slicks and lookalikes in polarimetric SAR images using CNN. *Sensors* 2017; **17**: 1837.
86. Guo H, Wei G and An J. Dark spot detection in SAR images of oil spill using Segnet. *Appl Sci* 2018; **8**: 2670.
87. Lee JS and Pottier E. *Polarimetric Radar Imaging: From Basics to Applications*. Boca Raton, FL: CRC Press, 2009.
88. Jones CE, Minchew B and Holt B *et al.* Studies of the Deepwater Horizon oil spill with the UAVSAR radar. *Geophys Monogr* 2013; **195**: 33–17.
89. Minchew B, Jones CE and Holt B. Polarimetric analysis of backscatter from the Deepwater Horizon oil spill using L-band synthetic aperture radar. *IEEE Trans Geosci Remote Sensing* 2012; **50**: 3812–30.
90. Maillard P, Clausi DA and Deng H. Operational map-guided classification of SAR sea ice imagery. *IEEE Trans Geosci Remote Sensing* 2005; **43**: 2940–51.
91. Fetterer F, Bertoia C and Ye JP. Multi-year ice concentration from Radarsat. In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS): Remote Sensing—a Scientific Vision for Sustainable Development 2017*. Singapore: IEEE, 402–3.
92. Lundhaug M. ERS SAR studies of sea ice signatures in the Pechora Sea and Kara Sea region. *Can J Remote Sens* 2002; **28**: 114–27.
93. Soh LK, Tsatsoulis C and Gineris D *et al.* ARKTOS: an intelligent system for SAR sea ice image classification. *IEEE Trans Geosci Remote Sensing* 2004; **42**: 229–48.
94. Zakhvatkina NY, Alexandrov VY and Johannessen OM *et al.* Classification of sea ice types in ENVISAT synthetic aperture radar images. *IEEE Trans Geosci Remote Sensing* 2012; **51**: 2587–600.
95. Leigh S, Wang Z and Clausi DA. Automated ice-water classification using dual polarization SAR satellite imagery. *IEEE Trans Geosci Remote Sensing* 2013; **52**: 5529–39.
96. Lang W, Wu J and Zhang X *et al.* Detection of ice types in the Eastern Weddell Sea by fusing L- and C-band SIR-C polarimetric quantities. *Int J Remote Sens* 2014; **35**: 6874–93.
97. Xu Y and Scott KA. Sea ice and open water classification of SAR imagery using cnn-based transfer learning. In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2017*. Fort Worth, TX: IEEE, 3262–4.
98. Gao Y, Gao F and Dong J *et al.* Transferred deep learning for sea ice change detection from synthetic-aperture radar images. *IEEE Geosci Remote Sensing Lett* 2019; **16**: 1655–9.
99. Celik T. Unsupervised change detection in satellite images using principal component analysis and k-means clustering. *IEEE Geosci Remote Sensing Lett* 2009; **6**: 772–6.
100. Gao F, Dong J and Li B *et al.* Change detection from synthetic aperture radar images based on neighborhood-based ratio and extreme learning machine. *J Appl Remote Sens* 2016; **10**: 046019.
101. Gao F, Dong J and Li B *et al.* Automatic change detection in synthetic aperture radar images based on PCANet. *IEEE Geosci Remote Sensing Lett* 2016; **13**: 1792–6.

102. Li J, Wang C and Wang S *et al.* Gaofen-3 sea ice detection based on deep learning. In: *2017 Progress in Electromagnetics Research Symposium-Fall (PIERS-FALL)*. Singapore: IEEE, 933–3.
103. Han Y, Gao Y and Zhang Y *et al.* Hyperspectral sea ice image classification based on the spectral-spatial-joint feature with deep learning. *Remote Sens* 2019; **11**: 2170.
104. Yi L, Zhang S and Yin Y. Influence of environmental hydro-meteorological conditions to *Enteromorpha prolifera* blooms in Yellow Sea, 2009. *Period Ocean Univ China* 2010; **40**: 15–8.
105. Hu C, Li D and Chen C *et al.* On the recurrent *Ulva prolifera* blooms in the Yellow Sea and East China Sea. *J Geophys Res* 2010; **115**: C05017.
106. van Tussenbroek BI, Arana HAH and Rodríguez-Martínez RE *et al.* Severe impacts of brown tides caused by *Sargassum* spp. on near-shore Caribbean sea-grass communities. *Mar Pollut Bull* 2017; **122**: 272–81.
107. Zhong S, Ding Y and Zhen L *et al.* Error analysis on *Enteromorpha prolifera* monitoring using MODIS data. *Remote Sens Inf* 2013; **28**: 38–42.
108. Hu C. A novel ocean color index to detect floating algae in the global oceans. *Remote Sens Environ* 2009; **113**: 2118–29.
109. Arellano-Verdejo J, Lazcano-Hernandez HE and Cabanillas-Terán N. ERISNet: deep neural network for *Sargassum* detection along the coastline of the Mexican Caribbean. *Peer J* 2019; **7**: e6842.
110. Eldhuset K. An automatic ship and ship wake detection system for spaceborne SAR images in coastal regions. *IEEE Trans Geosci Remote Sensing* 1996; **34**: 1010–9.
111. Vachon PW, Campbell J and Bjerkelund C *et al.* Ship detection by the RADARSAT SAR: validation of detection model predictions. *Can J Remote Sens* 1997; **23**: 48–59.
112. Ouchi K, Tamaki S and Yaguchi H *et al.* Ship detection based on coherence images derived from cross correlation of multilook SAR images. *IEEE Geosci Remote Sensing Lett* 2004; **1**: 184–7.
113. Iervolino P, Guida R and Whittaker P. A novel ship-detection technique for Sentinel-1 SAR data. In: *2015 IEEE 5th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*. Singapore: IEEE, 797–4.
114. Deng Z, Sun H and Zhou S *et al.* Learning deep ship detector in SAR images from scratch. *IEEE Trans Geosci Remote Sensing* 2019; **57**: 4021–39.
115. Liu L, Pan Z and Lei B. Learning a rotation invariant detector with rotatable bounding box. arXiv: 1711.09405.
116. An Q, Pan Z and Liu L *et al.* DRBox-v2: an improved detector with rotatable boxes for target detection in SAR images. *IEEE Trans Geosci Remote Sensing* 2019; **57**: 8333–49.
117. Lin TY, Dollár P and Girshick R *et al.* Feature pyramid networks for object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017*. Honolulu: IEEE, 2117–8.
118. Huang L, Liu B and Li B *et al.* OpenSARShip: a dataset dedicated to Sentinel-1 ship interpretation. *IEEE J Sel Top Appl Earth Observ Remote Sens* 2017; **11**: 195–208.
119. Villon S, Chaumont M and Subsol G *et al.* Coral reef fish detection and recognition in underwater videos by supervised machine learning: comparison between Deep Learning and HOG+SVM methods. In: *International Conference on Advanced Concepts for Intelligent Vision Systems 2016*. Lecce: Springer, 160–11.
120. Xu L, Bennamoun M and An S *et al.* Deep learning for marine species recognition. In: Balas VE, Roy SS and Sharma D *et al.* (eds.). *Handbook of Deep Learning Applications*. Basel: Springer, 2019, 129–45.
121. Marini S, Fanelli E and Sbragaglia V *et al.* Tracking fish abundance by underwater image recognition. *Sci Rep* 2018; **8**: 13748.
122. Saqib M, Khan SD and Sharma N *et al.* Real-time drone surveillance and population estimation of marine animals from aerial imagery. In: *International Conference on Image and Vision Computing New Zealand (IVCNZ) 2018*. Auckland: IEEE, 1–5.
123. Pedersen M, Bruslund Haurum J and Gade R *et al.* Detection of marine animals in a new underwater dataset with varying visibility. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops 2019*. Long Beach, CA: IEEE, 18–26.
124. Redmon J and Farhadi A. Yolov3: an incremental improvement. arXiv: 1804.02767.
125. Russakovsky O, Deng J and Su H *et al.* Imagenet large scale visual recognition challenge. *Int J Comput Vis* 2015; **115**: 211–52.
126. Xu Y, Li Y and Zhan Z *et al.* Morphology and phylogenetic analysis of two new deep-sea species of *Chrysogorgia* (Cnidaria, Octocorallia, Chrysogorgiidae) from Kocebu Guyot (Magellan seamounts) in the Pacific Ocean. *ZooKeys* 2019; **881**: 91–107.
127. Li Y, Zhan Z and Xu K. Morphology and molecular phylogeny of *Paragorgia rubra* sp. nov. (Cnidaria: Octocorallia), a new bubblegum coral species from a seamount in the tropical Western Pacific. *Chin J Ocean Limnol* 2017; **35**: 803–14.
128. Bewley M, Friedman A and Ferrari R *et al.* Australian sea-floor survey data, with images and expert annotations. *Sci Data* 2015; **2**: 150057.