Satellite remote sensing of pelagic *Sargassum* macroalgae: The power of high resolution and deep learning

Mengqiu Wang and Chuanmin Hu*

College of Marine Science, University of South Florida, 140 Seventh Avenue South, St. Petersburg, FL 33701, USA

*Corresponding author email: huc@usf.edu

Abstract In recent years, massive blooms of pelagic *Sargassum* have occurred in the Atlantic Ocean, Caribbean Sea, and Gulf of Mexico, and satellite imagery have been used operationally to monitor and track the blooms. However, limited by the coarse resolution and other confounding factors, there is often a data gap in nearshore waters, and the uncertainties in the estimated *Sargassum* abundance in offshore waters are also unclear. Higher-resolution satellite data may overcome these limitations, yet such a potential is hindered by the lack of reliable methods to accurately detect and quantify *Sargassum* in an automatic fashion. Here, we address this challenge by combining large quantities of **high-resolution satellite data** with **deep learning**. Specifically, data from the Multispectral Instrument (MSI, 10-20 m), Operational Land Imager (OLI, 30 m), WorldView-II (WV-2, 2 m), and PlanetScope/Dove (3 m) are used with a deep convolution neural network (DCNN) to extract *Sargassum* features and quantify *Sargassum* biomass density or areal coverage. By utilizing the U-net architecture and the pre-trained weights from the VGG16 model, the DCNN (i.e., the VGGUnet model) can extract *Sargassum* features while discarding other confusing features (waves, currents, phytoplankton blooms, clouds, cloud shadows, or striping noise). For *Sargassum* biomass estimated from OLI and MSI images, results indicate an accuracy of ~92% and 90%, respectively, when evaluated using images from the same sensor. When *Sargassum* areal coverage is estimated from WV-2 and Dove images, there is an accuracy of ~98% and 82%, respectively. When different sensors are cross-compared, OLI reveals ~30% more

Sargassum biomass than MODIS from 14 OLI images collected in the Caribbean Sea (path/row: 001/050) for their commonly viewed observable areas, and ~180% more *Sargassum* biomass than MSI (N = 15, path/row: 001/050); such differences appear systematic (R^2 = 0.98 and 0.73, respectively). Compared to the quasi-simultaneous MSI, OLI, and MODIS images, Dove shows higher *Sargassum* coverage. Higher-resolution sensors tend to observe more *Sargassum* because they can detect smaller-scale features that are missed by the coarser-resolution sensors, although the difference varies with time and location. The morphological characteristics of *Sargassum* features from these high-resolution data are also reported to facilitate management actions. The findings here not only fill the knowledge gaps and coverage gaps from previous studies, but more importantly pave the road toward operational monitoring and tracking *Sargassum* features in nearshore waters.

Keywords: Landsat-8, Sentinel-2, MSI, OLI, Worldview-2, Dove, MODIS, FAI, *Sargassum*, clouds, feature extraction, deep convolution neural network, deep learning, U-net, VGGUnet

Highlights

A deep learning *Sargassum* extraction model developed for high-resolution data

Model applicable to MSI, OLI, and 3-band Dove images with high accuracy

Uncertainties in coarse-resolution *Sargassum* abundance images quantified

Near-shore data gaps filled with high-resolution *Sargassum* abundance images

1. Introduction

During the past decade, the amount of pelagic *Sargassum* has increased significantly across the Atlantic Ocean (Wang et al., 2019). Consequently, coastal regions around the Caribbean Sea (CS), West Africa (WA), and Florida have experienced severe *Sargassum* beaching events (Smetacek & Zingone, 2013; Hu et al. 2016a; Gower & King, 2019a; van Tussenbroek et al., 2017; Rodríguez-Martínez et al., 2019; Chávez et al., 2020). Massive *Sargassum* deposition on the beaches has induced various environmental, ecological, and human health problems, and negatively impacted local economies (Siuda, Schell, & Goodwin 2016; Langin, 2018; Rodríguez-Martínez et al., 2019; van Tussenbroek et al., 2017).

Satellite remote sensing provides timely information for monitoring and tracking of *Sargassum*, and thus is a useful tool to help resource managers make decisions and develop mitigation strategies, and may also help site suitability assessment for *Sargassum* farming or harvesting (Webster & Linton, 2013; Hu et al., 2016; Wang & Hu, 2017; Zheng et al., 2018; Xing et al., 2018; Xing et al., 2019; Bach et al., 2021). Currently, coarse-resolution sensors including the Moderate Resolution Imaging Spectroradiometer (MODIS), Visible Infrared Imaging Radiometer Suite (VIIRS), MEdium Resolution Imaging Spectrometer (MERIS), and Ocean Land Colour Instrument (OLCI) have been successfully utilized to observe the large-scale *Sargassum* distributions across the Atlantic Ocean (Wang & Hu, 2016, Wang et al., 2018, Wang et al., 2019, Gower & King, 2019b). Correspondingly, a satellite-based near real-time *Sargassum* Watch System (SaWS) has been established to use both MODIS and VIIRS imagery to monitor *Sargassum* distributions and to predict *Sargassum* transport in the CS (Hu et al., 2016a; Wang & Hu, 2016; Wang et al., 2018) (Fig. 1a).

However, measurements derived from these coarse-resolution sensors often suffer from several limitations. First, **uncertainties in the *Sargassum* estimates are often unclear**. *Sargassum* in the ocean can take the form of clumps, mats, or rafts, often smaller than a pixel size (Ody et al., 2019). Each sensor has its own lower detection limit. For example, with a signal-to-noise ratio (SNR) of

200:1, Hu et al. (2015) estimated a subpixel detection limit of about 1% of a pixel size. From MODIS 1-km observations, the lower detection limit was estimated to be 0.2% of a pixel size (Wang & Hu, 2016), or 2000 m^2. It is unclear how much *Sargassum* these coarse-resolution sensors may "miss" due to such lower detection limits, for example in the weekly *Sargassum* density images (Fig. 1b). In previous studies, high-resolution data have been used to evaluate macroalgae abundance estimated from MODIS (Hu et al, 2016b; Cui et al. 2018; Hu et al., 2019; Wang et al., 2021), but only a very limited number of images were used in these comparisons. Moreover, there are **no valid MODIS or VIIRS observations in nearshore waters** (Fig. 1b). As explained in Wang and Hu (2020), the data quality of the coarse-resolution pixels is compromised in coastal waters due to interference of the shallow-water bottom, high amounts of suspended particles, or land adjacency effects. Therefore, in the SaWS, pixels within 30 km of shoreline are often masked to avoid false positives (Hu et al., 2016a, Wang & Hu, 2020). This lack of data in nearshore waters greatly hinders management actions.

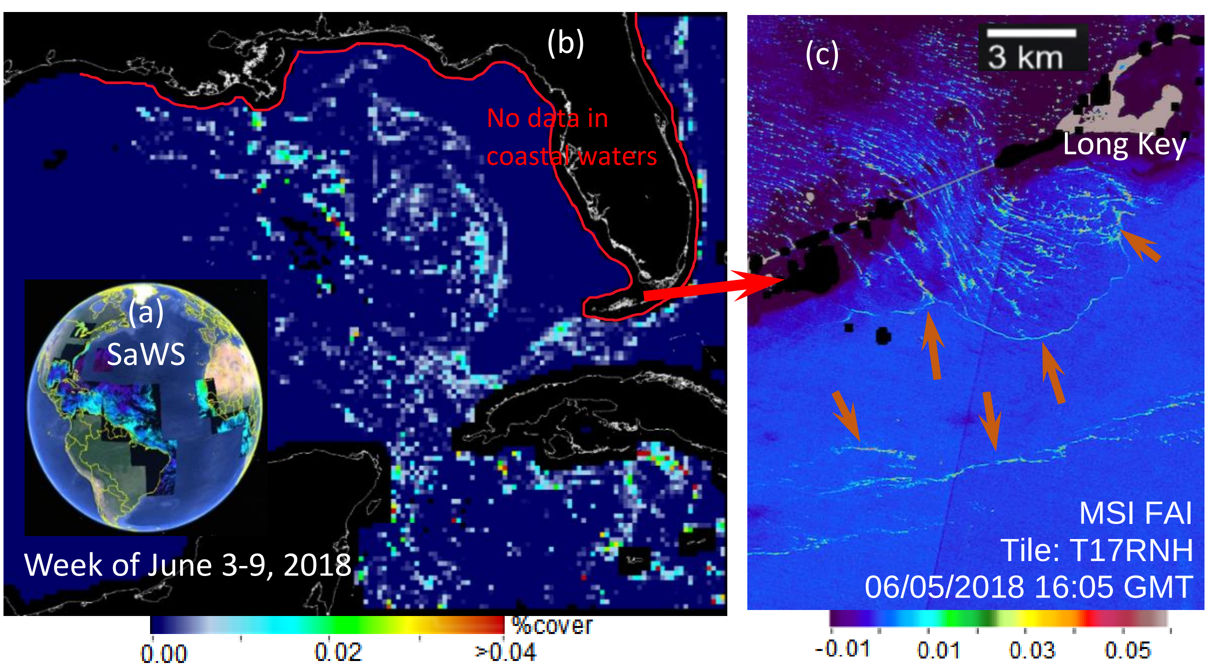


Fig. 1. (a) SaWS domain shown in Google Earth (optics.marine.usf.edu/projects/saws.html), which covers the entire Caribbean Sea and Gulf of Mexico. (b) A weekly *Sargassum* density map provided by SaWS shows *Sargassum* transport from the Caribbean to the Florida Straits. A value of 0.02 indicates 0.02%. **Note that there is no data coverage in nearshore waters** (30 km within shoreline) as indicated by the red outline**.** (c) MSI FAI image shows excellent coverage for coastal and nearshore waters around Long Key in the Florida Keys. The image slicks (annotated by the arrows) represent *Sargassum* rafts.

To overcome these limitations, various **high-resolution sensors** should be utilized**.** For example, the 10–30 m resolution Multispectral Instrument (MSI) and Operational Land Imager (OLI) sensors carried by the Sentinel-2 and Landsat-8 satellites are equipped with spectral bands to detect the enhanced Near Infrared (NIR) reflectance caused by floating vegetations including *Sargassum*. Fig. 1c shows a MSI Floating Algae Index (FAI, Hu, 2009) image around Long Key of the Florida Keys, where *Sargassum* rafts can be clearly visualized as image slicks. Such observations are not available from the coarse-resolution data in Fig. 1b for the same period. Figs. 2a & b show more examples of *Sargassum* features in the MSI and OLI FAI images, respectively. In addition, commercial satellites such as the Worldview series (Fig. 2c), PlanetScope (Fig. 2d), RapidEye, and SkySat also provide high-resolution data that can be used to detect the small floating macroalgae features. In particular, the PlanetScope constellation (Dove) provides daily observations at 3-m resolution for many coastal areas, thus representing an excellent data source for near real-time applications. As an example, the 3-m resolution Dove image in Fig. 2d shows many small brownish *Sargassum* features that can be visualized without the NIR wavelengths.

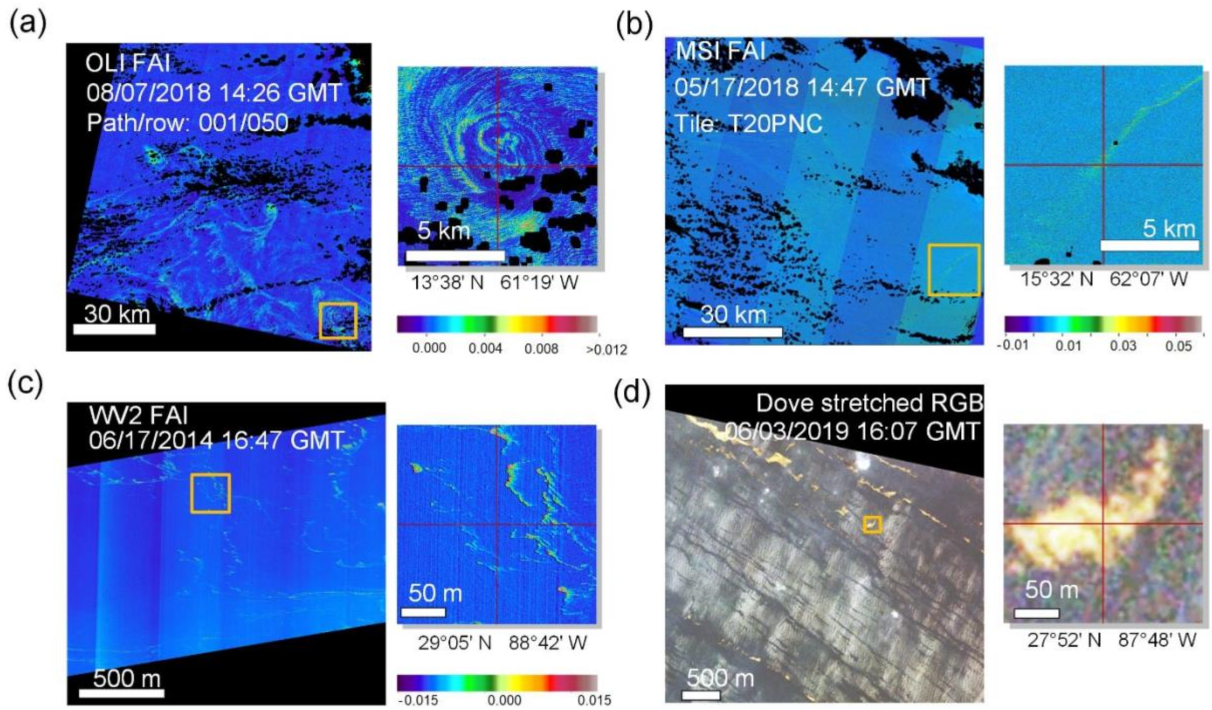


Fig. 2. ***Sargassum* features in high-resolution satellite images.** (a) – (c): OLI, MSI, and WV-2 FAI images, respectively; (d): Dove stretched RGB images. Without contrast enhancement through Gaussian stretching, most of these *Sargassum* features are invisible. Due to the space limit, only small portions (outlined in squares whose center coordinates are labeled below the sub-images) can be displayed in full resolution. Black color represents no observations due to either clouds or no data coverage. Note that there are various noises (including stripes and wave-induced glitters) and large-scale background variations in all images.

While these high-resolution sensors are designed primarily for land-based applications, some measurements are also taken over the ocean (Hedley et al., 2018); however, *Sargassum* detection often suffers from confusing features induced by clouds, surface waves, or variable image background due to sensor artifacts or changes in water's optical properties (Wang & Hu, 2020). Some of these confusing features can be spectrally similar to weak *Sargassum* features (Fig. S1). To make things worse, different sensors may have different noise characteristics (Figs. S2-S5). Thus, a reliable algorithm to extract *Sargassum* features automatically from images of an individual sensor, not to mention a unified algorithm applicable to all high-resolution sensors, is lacking. Because of these technical difficulties, in the *Sargassum* Early Advisory System (SEAS), visual interpretation and manual delineation are often required to locate the *Sargassum* slicks in the Landsat imagery to predict potential beaching events (Webster & Linton, 2013).

Although a denoising and feature extraction method has been developed for MSI applications (Wang & Hu, 2020), it is only designed to work **on FAI images** that require at least one spectral band in the red and two spectral bands in the NIR or SWIR wavelengths. Such a requirement cannot be met by the 3-band data from the PlanetScope constellations. In addition, the threshold-based image segmentation method relies on the **accurate estimation of the image background variations,** and may need tuning to meet individual sensor needs.

On the other hand, **deep learning-based techniques** have shown great potential in analyzing features from ocean remote sensing images, with many successful applications for ship detection and macroalgae extraction (Li et al, 2020; Ma et al., 2019; Hordiiuk, Oliinyk, Hnatushenko, & Maksymov, 2019; Arellano-Verdejo, Lazcano-Hernandez, & Cabanillas-Terán, 2019, Wang S. et al., 2019). For example, the ERISNet, a one-dimensional DCNN framework proposed by Arellano-Verdejo et al. (2019), shows good performance in *Sargassum* extraction in coastal waters around Mexico. In Wang S. et al. (2019), AlexNet is applied to classify the macroalgae patches from the Unmanned Aerial Vehicle (UAV) images. Comparatively, because the classic U-net model (Ronneberger, Fischer, & Brox, 2015) has a more complex and efficient network structure (which has shown great image segmentation performance on both biomedical and remote sensing images), it should also have the capacity to detect and extract *Sargassum* from high-resolution images.

Therefore, by using the U-net model, the objective of this paper is to design and develop a unified approach to extract *Sargassum* features and quantify *Sargassum* abundance from multi-sensor high-resolution imagery, with the ultimate goal of addressing the following questions: 1) how much *Sargassum* have coarse-resolution (e.g., MODIS) observations missed? 2) how to fill the data gaps in nearshore waters?

To achieve these goals, this paper is structured as follows: the *Sargassum* quantification workflow and the details of the VGGUnet model (the DCNN used for *Sargassum* extraction) are first described, followed by a performance evaluation on the MSI, OLI, WV-2, and Dove datasets. The *Sargassum* biomasses/coverages quantified from these high-resolution images are compared with the concurrent MODIS measurements to establish empirical relationships between these sensors. The *Sargassum* morphology measured from the OLI, MSI, and Dove images are analyzed. Lastly, the limitations of satellite remote sensing of *Sargassum*, the strengths and weaknesses of this approach, and operational considerations for near real-time *Sargassum* monitoring are discussed.

2. Data and Methods

2.1 Data preparations

2.1.1 Sentinel-2 and Landsat-8 data

Fifty-three Sentinel-2 MSI and twenty-one Landsat-8 OLI Level-1C (top-of-Atmosphere (TOA) reflectance) images collected near the Lesser Antilles Islands and Gulf of Mexico (GOM) in 2018 and 2019 were downloaded from the USGS earth explorer https://earthexplorer.usgs.gov/, and processed with ACOLITE (Vanhellemont & Ruddick, 2015, version 20190326) to Rayleigh-corrected reflectance (R_{rc}, unitless) at 10-m and 30-m resolution, respectively. Using the multispectral R_{rc} data, the FAI products were generated to quantify the enhanced reflectance of *Sargassum* in the near-infrared (NIR) wavelengths by comparing to the nearby RED and ShortWave-Infrared (SWIR) bands using the following equation:

$$FAI = R_{rc,NIR} - R'_{rc,NIR}$$

$$R'_{rc,NIR} = R_{rc,RED} + (R_{rc,SWIR} - R_{rc,RED}) \times (\lambda_{NIR} - \lambda_{RED}) / (\lambda_{SWIR} - \lambda_{RED}) \quad (1)$$

where $\lambda_{RED} = 665$ nm, $\lambda_{NIR} = 865$ nm, and $\lambda_{SWIR} = 1610$ nm were selected for Sentinel-2 MSI data, while $\lambda_{RED} = 655$ nm, $\lambda_{NIR} = 865$ nm, and $\lambda_{SWIR} = 1610$ nm were selected for Landsat-8 OLI data.

On MSI FAI images, the pixels with large R_{rc1610} (> 0.10) were pre-masked to exclude the land and bright cloud pixels and treated as invalid observations (Eq. (2)). Similar thresholds were also applied to mask the OLI FAI images before *Sargassum* extraction.

$$R_{rc1610} > 0.1,\ R_{rc442} > 0.1,\ \text{and}\ R_{rc560} > 0.1 \quad (2)$$

2.1.2 Worldview-2 data

Four Worldview-2 (WV-2) images collected in the northern GOM during 2014 to 2015 containing partial *Sargassum* coverage were acquired from DigitalGlobe. The data were processed to generate TOA reflectance, and the FAI products were generated using the TOA reflectance centered on 659nm, 833nm, and 949nm. Fig. 2c shows an example of the *Sargassum* features observed on the WV-2 FAI images. Three images were used for model training and one image was selected for validation.

2.1.3 Dove data

A total of 1, 4,567, and 7,457 three-band Dove images collected on 1 June 2019, 3 June 2019, and 5 June 2018, respectively, over the GOM were acquired from Planet Lab to test the applicability of the *Sargassum* extraction method. The one image on 1 June 2019 corresponded to field measurements on the same day, and therefore was selected in training dataset. On the other two days, all Dove images over the GOM available at Planet Lab were acquired and used in this study. Considering the difficulties in conducting accurate atmospheric correction due to lack of at least two NIR bands and due to variable sun glint and sky glint (Wicaksono and Lazuard, 2018), the TOA radiance data were directly used to detect *Sargassum*. Because of similar solar and viewing angles across the narrow-swath (< 30 km) Dove images, Rayleigh correction is equivalent to removing a constant from all pixels, and therefore was not performed. The four-band Dove data which contains the NIR wavelengths were mostly unavailable in the open water area within the GOM, therefore only the three-band RGB data were used in this paper. Table 1 summarizes the details of the high-resolution images analyzed in this study.

Table 1: High-resolution satellite images used for detecting and quantifying *Sargassum* in this study. Here Dove data only refer to the three-band data because they have daily coverage over the GOM, while four-band data (with the 4-th band in the NIR wavelength) cover coastal waters only.

	Spatial resolution	**Revisit time**	**NIR bands**	**Image location**	**Number of images used**	**Model input**
MSI	10-m/20-m	5 days	Yes	Near the Lesser Antilles Islands; Eastern GOM	53	FAI
OLI	30-m	16 days	Yes	Near the Lesser Antilles Islands (path/row: 001/050; 002/049); Northern GOM (path/row: 021/040; 021/041)	21	FAI
Dove	3-m	Daily	No	GOM	12025	RGB
WV-2	~2-m	Irregular	Yes	Northern GOM	4	FAI

2.1.4 MODIS data

To estimate the amount of *Sargassum* missed by coarse-resolution sensors, MODIS data collected in the GOM on 3 June 2019 and 5 June 2018 and in the Central West Atlantic in 2018 were processed to compare with quasi-simultaneous and co-located MSI, OLI, and Dove observations. MODISA and MODIST Level-0 data were obtained from the U.S. National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (http://oceancolor.gsfc.nasa.gov), and processed to generate R_{rc} data using SeaDAS software (version 7.5). The corresponding MODIS Alternative FAI (AFAI) images were generated using R_{rc} data centered at 667 nm, 748 nm, and

869 nm (Eq. (1)). The *Sargassum*-containing pixels were extracted, and the fractional areal coverages were quantified using a linear unmixing method (Wang and Hu, 2016). These area coverages were converted to biomass densities using the biomass model proposed in Wang et al. (2018).

2.2 *Sargassum* extraction and quantification workflow

In this study, *Sargassum* extraction and quantification follow a straightforward workflow: **First**, data under cloudy conditions and other unfavorable observing conditions were treated as no observations. **Then**, the *Sargassum*-containing pixels were extracted using the VGGUnet model trained for the specific data types (section 2.2.2). **Finally**, the corresponding biomass densities/areal coverages were quantified from all *Sargassum*-containing pixels.

Fig. 3 illustrates the major workflow applied to the Landsat-8 OLI and Sentinel-2 MSI images. For WV-2 and Dove images, *Sargassum* extraction was achieved using the same VGGUnet model, but in the final step *Sargassum*-containing pixels were assumed to have 100% subpixel areal coverage. Cloud masking was not considered for WV-2 data, while for Dove data blue-band radiance > 17 $W \cdot sr^{-1} \cdot m^{-2}$ was selected to mask thick clouds.

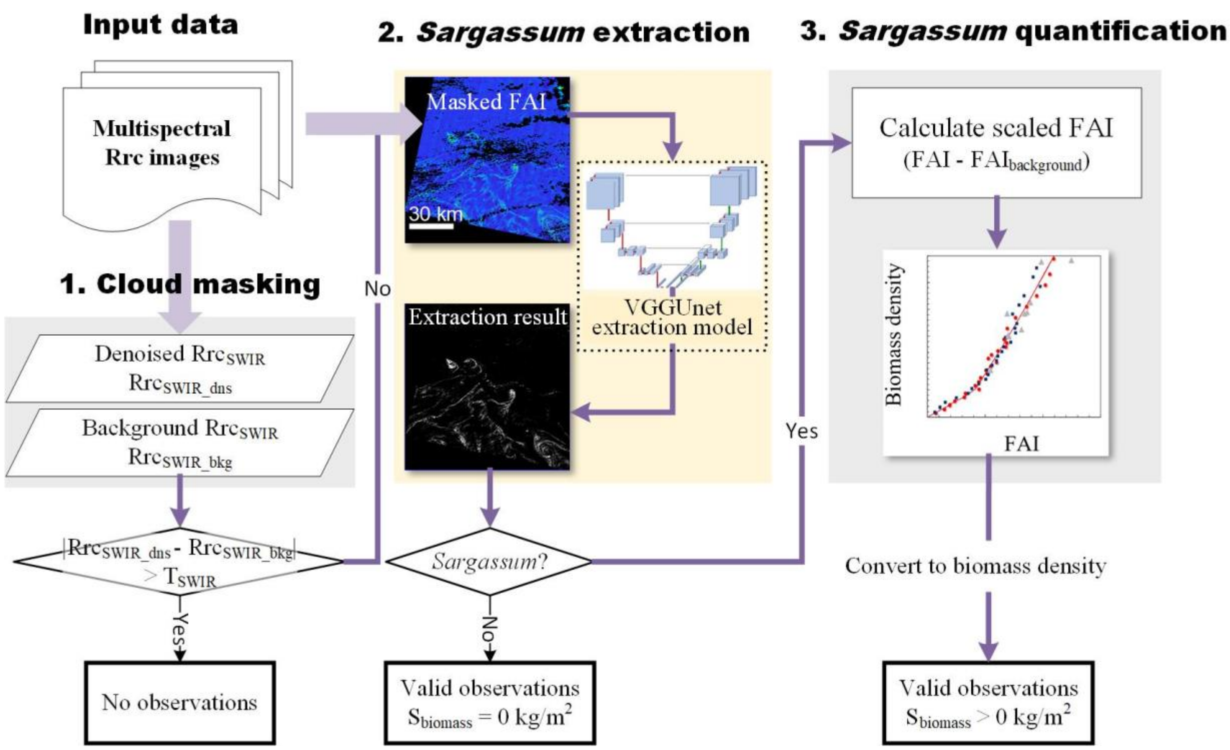


Fig. 3. Workflow for the automatic *Sargassum* detection and quantification using MSI/OLI images. The methods for cloud masking and *Sargassum* extraction are explained in section 2.2.1 and section 2.2.2, respectively. *Sargassum* extraction was realized through a VGGUnet model trained for the specific dataset. Pixels of valid observations but with no *Sargassum* detected were assigned with 0.0 kg/m^2 biomass density, while pixels with *Sargassum* detected were assigned with the biomass density values calculated using the pixels' FAI values and a field-based FAI-biomass density model (Wang et al., 2018; Wang & Hu, 2020).

2.2.1 Preprocessing to mask pixels under cloudy conditions or other unfavorable observing conditions

Compared to background seawater, cloud pixels also show enhanced signals on FAI images and thus need to be masked before applying the *Sargassum* extraction process. Because clouds normally show higher reflectance in the SWIR wavelengths, a simple threshold can remove most thick clouds (Eq. (2)). However, this preliminary mask cannot identify thin clouds, and a unified threshold may over-mask valid water observations under strong sun glint. **For MSI and OLI data**, a H_SWIR cloud mask proposed in Wang and Hu (2020) was applied to mask the cloud-contaminated pixels. Instead of directly applying the single threshold over the entire image, the

H_SWIR cloud mask conducts the image segmentation after estimating the scaled reflectance by subtracting background reflectance (Eq. (3)).

$$Rrc_{SWIR_{dns}} - Rrc_{SWIR_{bkg}} > T_{SWIR} \tag{3}$$

where $Rrc_{SWIR_{dns}}$ is the denoised Rrc_{SWIR} with Gaussian filtering, and $Rrc_{SWIR_{bkg}}$ is the estimated background value of the Rrc_{SWIR}. The parameters selected for the MSI data have been discussed in Wang and Hu (2020). For the OLI data, the SWIR band centered at 1609 nm was applied to generate a similar H_SWIR cloud mask, which has demonstrated satisfactory performance on the tested OLI images through visual inspection.

On Dove images, cloud features are highly variable, making it challenging to effectively identify them. Additionally, even under moderately thick clouds, it is still possible to determine the *Sargassum* presence. Therefore, only those pixels with blue radiance greater than 17 $W \cdot sr^{-1} \cdot m^{-2}$ were masked as invalid observations. **The WV-2 images** used in this study are mostly cloud free, and cloud masking was not considered.

2.2.2 VGGUnet model for *Sargassum* extraction

2.2.2.1 Model structure

In this study, a deep learning framework (the VGGUnet model) combined with a U-net structure and the VGG-16 encoder was designed for *Sargassum* extraction from high-resolution satellite images. The U-net architecture was first proposed by Ronneberger, Fischer, & Brox (2015) for image segmentation on biomedical images. This unique architecture is able to capture context, as well as to precisely locate targeted features, thus has been tested for feature detection tasks on remote sensing images and achieved superior performance than traditional approaches (Iglovikov, Mushinskiy, & Osin, 2017; Li et al., 2020). It was also found that using the pre-trained encoder optimized on the Images-Net dataset can further improve the segmentation performance (Deng et al. 2009; Iglovikov, & Shvets, 2018). Therefore, the pre-trained weights from the VGG16 model (Simonyan, & Zisserman, 2014, see the purple arrows in Fig. 4) were used in the VGGUnet model. The detailed structure is introduced in Fig. 4. The total number of parameters was 35,120,069, of which 20,397,571 were trainable and 14,722,498 were non-trainable (adopted from the VGG16 model).

The input of the VGGUnet model can be either single-band or multi-spectra images. Because *Sargassum* shows enhanced signals and distinctive spatial patterns on the FAI images, FAI images were selected as the model input to determine the *Sargassum* locations on MSI, OLI, and WV-2 images. For Dove, the three-band RGB images were used as the model input due to the lack of NIR bands. The model output are the extracted feature pixels, which refer to the *Sargassum*-containing pixels in this study.

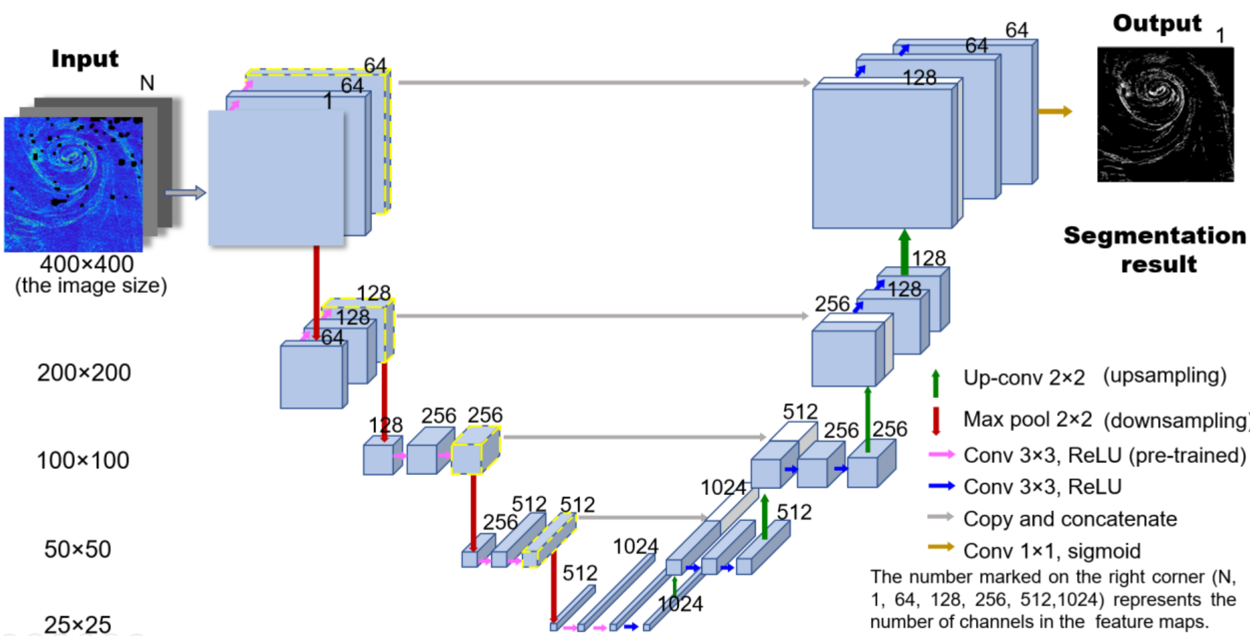


Fig. 4. The structure of the VGGUnet model, adapted from the U-net with VGG16 encoders (Simonyan, & Zisserman, 2014). It consists of a contracting path (left side, taken from the VGG16 model) and an

expansive path (right side, following the classic U-net design). Each blue cube represents a multichannel feature map. The white cubes represent the copied feature maps (indicated by the yellow dashed lines and the gray arrows). The image size in each of the 5 rows is marked in the first column (e.g., 100 × 100). The number of channels of each feature map is annotated on the upper right corner. For example, the N marked on top of the input image means that there are N spectral bands in the input image and the 1 marked on top of the output image means that there is only one channel in the output layer. Batch normalization was applied to normalize each convolutional block (Ioffe & Szegedy, 2015). The Rectified Linear Unit (ReLU) was used as the primary activation function. The sigmoid activation function was selected in the final output layer to determine the segmentation results. Note that the model input is flexible, where multispectral data can be used.

2.2.2.2 Model training and prediction

To optimize the VGGUnet model for the specific feature extraction tasks, the corresponding training datasets (consisting of the input images and the segmentation results) were prepared. Here, the ground "truth" of the *Sargassum* extraction results were generated using a semi-automatic IDL feature extraction Graphic User Interface (GUI, Wang & Hu, 2015). A total of 3,289 sub-images were prepared for MSI FAI images (from 14 MSI image tiles), 1,444 sub-images were selected for OLI FAI images (from 4 OLI image scenes), 682 sub-images were selected for WV-2 FAI images (from 3 WV-2 images), and 1791 sub-images were prepared on Dove RGB images (from 12 Dove images). These extraction results were cut into 400 × 400 sub-images to train the extraction model. Because there are already sufficient training images prepared for each sensor under various conditions, data augmentation techniques were not used.

During model optimization, the Jaccard Index (JI, Eq. (4)) was monitored to determine the similarity between the model outputs and the training data.

$$JI(y_{pred}, y_{true}) = \frac{1}{n}\sum_{i=1}^{n}\frac{y_{pred}\cdot y_{true}+smooth}{y_{pred}+y_{true}-y_{pred}\cdot y_{true}+smooth} \quad (4)$$

where y_{pred} is the continuous prediction probability values ($y_{pred} \in [0, 1]$) and y_{true} is the binary values from the ground-truth results ($y_{true} \in \{0, 1\}$). The smooth term is 1. Then, the degree of prediction inconsistency can be defined as the Jaccard Distance shown in Eq. (5).

$$JD(y_{pred}, y_{true}) = -\log JI(y_{pred}, y_{true}) \tag{5}$$

Because image segmentation is essentially a one class classification problem, the loss function L was defined as the JI after adding the binary cross-entropy term H (Eqs. (6-7)).

$$H(y_{pred}, y_{true}) = \frac{1}{n}\sum_{i=1}^{n}\left(y_{pred}\log\left(y_{true}\right) + (1 - y_{pred})\log\left(1 - y_{true}\right) \right) \tag{6}$$

$$L(y_{pred}, y_{true}) = H(y_{pred}, y_{true}) + JD(y_{pred}, y_{true}) \tag{7}$$

The Adaptive Moment estimation (adam) optimizer (Kingma & Ba, 2014) was applied for model optimization. The initial learning rate was 0.001. When the loss function failed to improve after two consecutive epochs, the learning rate would then be reduced by 20% for finer tuning. In our experiments, all models were trained for 200-300 epochs as stable performance was often achieved by that time, with high JI values in the training and validation dataset. Table 2 summarizes the estimated training time used on each dataset. In all four cases, the models can be effectively optimized within 24 hours.

Table. 2. The **approximate training time** of the VGGUnet model used on the high-resolution training images. Here all the sub-images are 400 × 400 pixels. The number in the bracket indicates the number of original images that these sub-images were selected from. In this paper, the experiments were conducted on the same PC with Intel(R) Core(TM) i9-9900 CPU @ 3.30GHz and a Nvidia GeForce RTX 2080 Ti GPU. Here, the batch size of 6 was used in model training due to limited memory availability.

Data	**MSI**	**OLI**	**WV-2**	**Dove**
Number of sub-images selected for training	3289 (14)	1444 (4)	682 (3)	1791 (12)
Batch size	6	6	6	6
Number of epochs trained	200	200	300	300
Average running time per epoch	257s	104s	48s	102s
Model training time	14.3 hours	5.8 hours	4.0 hours	8.5 hours

To use the VGGUnet model for *Sargassum* detection, input large satellite images (FAI or RGB) were cut into 416×416 sub-images. As the prediction accuracy could decrease along image edges (the boundary effect, Iglovikov, Mushinskiy, & Osin, 2017), these sub-images were prepared with redundant edges (8 pixels outward on four directions) and only the prediction results from the image center (with 400×400 pixels) were merged back to generate the final extraction results.

2.2.3 *Sargassum* biomass density quantification

To quantify *Sargassum* biomass density, the background FAI values were first estimated to account for the reflectance variations of the background water. The background FAI values were then subtracted to calculate the scaled FAI to estimate the corresponding biomass density. This is the same method as discussed in Wang and Hu (2020) for the MSI data. For the OLI data, the background estimation parameters and the FAI-biomass models were similarly applied, through an iterative median filtering (with a 200 × 200 window) and the following FAI-biomass model (Eq. (8) and Fig. 5b).

$$y = 22.89x \qquad (0 < x \leq 0.05)$$

$$y = 57.42\,(1.18x - 0.06)^2 + 36.00(1.18x - 0.06) + 1.17 \qquad (x > 0.05) \qquad (8)$$

where x is the OLI FAI values and y is the modeled *Sargassum* biomass density (kg/m^2).

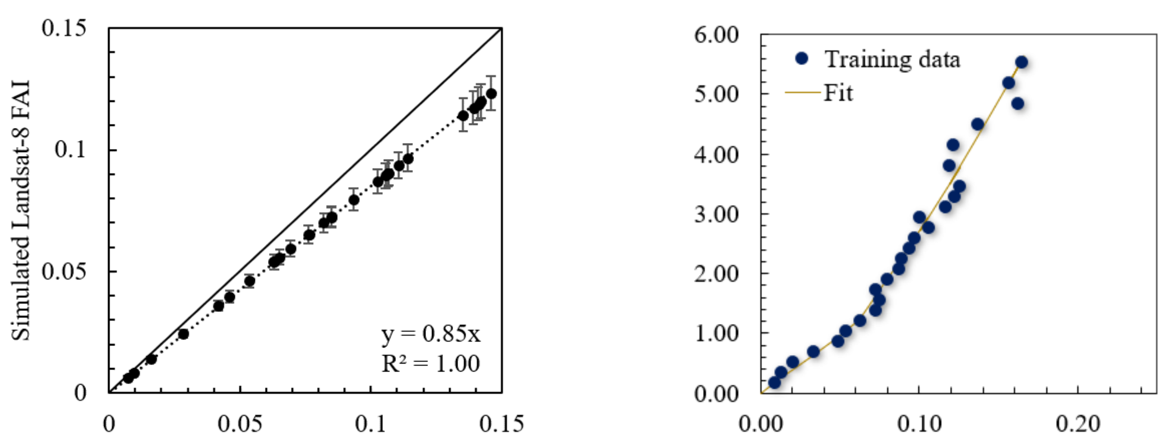


Fig. 5. (a) Comparison between *in situ* OLI FAI and simulated OLI FAI, with the latter being simulated by propagating *in situ* OLI FAI to top of atmosphere with aerosol optical thickness at 869 nm, $\tau_a(869) = 0.10$, averaged under different aerosol types and viewing geometry. Here, *in situ* OLI FAI stands for the FAI value calculated from the field-measured *Sargassum* spectra using the spectral response functions of the corresponding OLI bands. The solid line is the 1:1 line and the dashed line is the fitted line. The standard deviations of the simulated FAI are indicated by the vertical error bars. (b) *Sargassum* biomass density (kg/m^2) versus *in situ* OLI FAI.

2.2.4 *Sargassum* areal coverage quantification

Considering the high spatial resolution of Dove and WV-2 and the difficulties of conducting accurate biomass quantification, all the *Sargassum*-containing pixels extracted from these two sensors were assigned 100% subpixel *Sargassum* areal coverage. For example, on the 3-m resolution Dove images, each extracted *Sargassum*-containing pixel was assumed to have 9 m^2 of *Sargassum*.

To compare with the Dove-derived *Sargassum* measurements, the *Sargassum* areal coverages derived from OLI and MSI were quantified through linear unmixing using a full-coverage threshold. Those pixels with biomass densities lower than the threshold were linearly unmixed to calculate the fractional coverage, while pixels with higher biomass densities were treated to have 100% *Sargassum* coverage (i.e. 900 m^2 for a 30-m OLI *Sargassum*-containing pixel and 100 m^2 for a 10-m MSI *Sargassum*-containing pixels). The full coverage thresholds were selected to be the biomass densities when FAI equals to 0.05 (the turning point of changing from linear to nonlinear relationships in the FAI-biomass model), and the values for the Sentinel-2A MSI, Sentinel-2B MSI, and Landsat-8 OLI data are 0.96, 1.24, and 1.17 kg/m^2, respectively. For MODIS data, the areal coverages were estimated using the method described in Wang and Hu (2016). When comparing Dove with coarser-resolution sensors, because only the dense *Sargassum* mats can show up in the high-resolution RGB images (Fig. S6), we did not perform linear unmixing but assumed that every extracted Dove pixel had full *Sargassum* coverage.

3. Results

3.1 *Sargassum* extraction performance from high-resolution satellite images

To evaluate the extraction accuracy, the proposed approach was tested on a separate group of representative MSI, OLI, Dove, and WV-2 images. The extraction results were then compared with the manually extracted "ground truth" features (Wang & Hu, 2015) to generate the corresponding F1 score (Chinchor & Sundheim, 1993). Fig. 6 illustrates the *Sargassum* extraction results from the high-resolution satellite images listed in Fig. 2. From visual inspection, satisfactory performance was achieved: no apparent noise signals were misidentified as *Sargassum* features (i.e., low false positives), and the *Sargassum*-containing pixels were mostly detected (i.e., low false negatives).

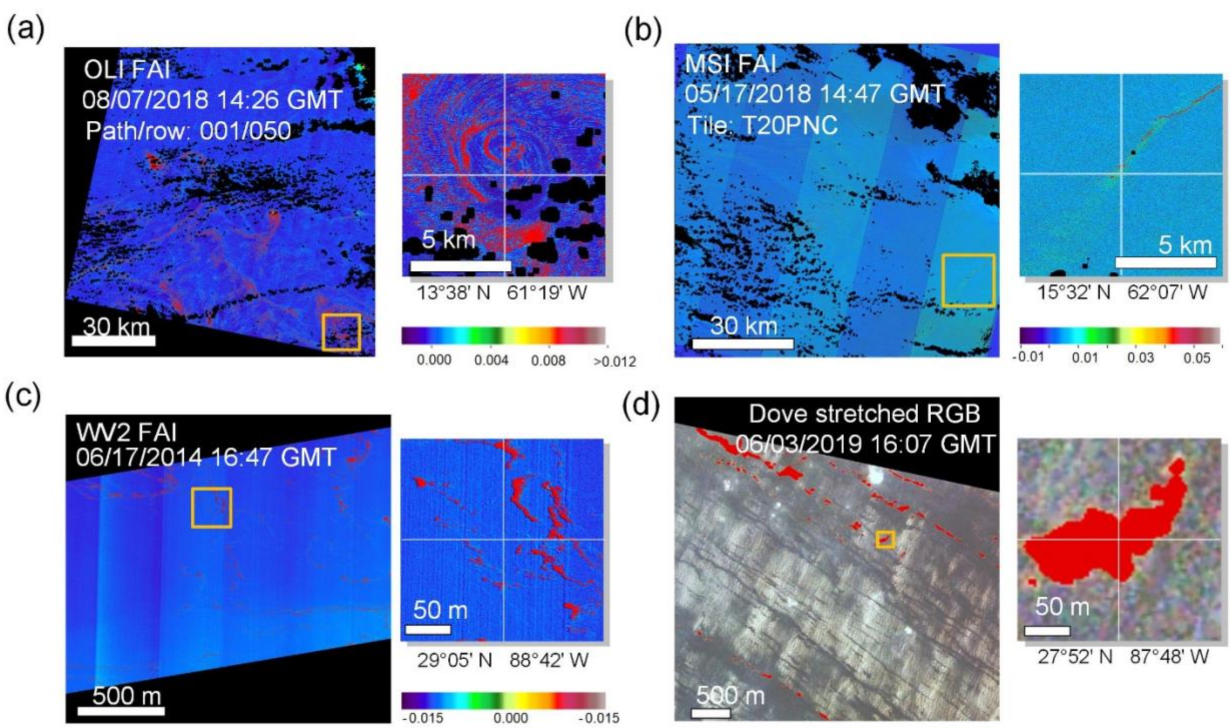


Fig. 6. *Sargassum* extraction from high-resolution satellite images using the VGGUnet model. These are the same images as shown in Fig. 2. (a) – (c): OLI, MSI, and WV-2 FAI images, respectively. (d) Dove stretched RGB images. The extracted *Sargassum*-containing pixels (colored in red) are overlaid on the images.

The extraction accuracy from individual sensors is listed in Table 3, including their false positive rates, false negative rates, precision, recall, and F1 score. Because the "ground truth" was not obtained from field measurements but generated from the manual work from the validation images, such the accuracy should be regarded as self-consistency accuracy.

On the MSI FAI images, the overall *Sargassum* extraction accuracy, after weighted by the biomass density, is ~90%. This is higher than the accuracy of the previous TNRD-based *Sargassum* extraction method (86%, Wang & Hu, 2020). Most of detection errors (either false positives or false negatives) are from pixels of relatively low biomass densities. The precision and recall rates are both > 85%, suggesting that most of the *Sargassum*-containing pixels can be accurately detected, and most detected candidate pixels contain *Sargassum*. Comparatively, small *Sargassum* patches are relatively harder to identify than larger *Sargassum* patches, but most features can still be effectively detected (Table S1).

On most OLI FAI images with large *Sargassum* coverages, the extraction accuracy is > 95% in terms of *Sargassum* biomass densities. The precision and recall rates are both higher than those from the MSI FAI images. The higher accuracy is likely due to the larger pixel size and less noise interference (such as wave glitters) than found in MSI FAI images.

Due to the higher spatial resolution and larger image size, for WV-2 FAI images and Dove RGB images, only a limited number of images were selected to evaluate the extraction accuracy. The areal coverage (as opposed to biomass density) was used to evaluate the accuracy. Table 3 shows that the accuracy for WV-2 is almost perfect (F1 score = 0.98). Even with 3 spectral bands in the visible wavelengths, Dove images still show promising performance, with F1 score greater than 0.8.

Overall, when evaluated using images collected by the same sensor, the approach shows F1-score of ≥ 0.90 except for the 3-band Dove images. Even for these images without the NIR bands, the F1-score is still 0.82, suggesting that the approach may be used to extract *Sargassum* features automatically.

Table 3. *Sargassum* extraction accuracy on MSI, OLI, WV-2, and Dove images using the VGGUnet model. For MSI and OLI, the biomass density (weighted by sub-pixel coverage) was compared between model results and "ground truth" images derived from the same sensor. For WV-2 and Dove, every detected *Sargassum* pixel is assumed to have 100% subpixel coverage in such comparisons. The accuracy can only be higher if each pixel is weighted by its sub-pixel coverage, as larger *Sargassum* feature are more likely to be accurately identified. The number of images in the table means number of original images, not the 400 × 400 sub-images.

	Number of images	Mean % of valid observations	False positive	False negative	Precision	Recall	F1 score
MSI	10	77%	0.05	0.15	0.95	0.85	**0.90**
OLI	8	50%	0.06	0.11	0.94	0.90	**0.92**
Dove	2	57%	0.38	0.04	0.72	0.96	0.82
WV-2	1	100%	0.01	0.04	0.99	0.96	0.98

3.2 Comparison of *Sargassum* amount estimated from Dove, MSI, OLI, and MODIS

One fundamental question on *Sargassum* detection using coarse-resolution satellite sensors such as MODIS is how much *Sargassum* may be missed. Now with the known accuracy of the proposed approach in extracting *Sargassum* features from multiple high-resolution sensors (Dove, MSI, and OLI) and with the availability of a large quantity of high-resolution images, this question may be addressed.

As shown in Fig. 7, the 3-m Dove images have much higher daily coverage over the GOM than other high-resolution sensors. Indeed, the PlanetScope constellation provides the only data source to cover the entire GOM nearly every day at 3-m resolution. Furthermore, nearly all *Sargassum* features extracted from MSI and OLI images are clearly revealed in the corresponding Dove images (Fig. 8), therefore, *Sargassum* extraction results from the 12,024 Dove images collected over the GOM on 3 June 2019 and 5 June 2018 were used as the truth to evaluate the extraction uncertainties from MODIS, MSI, and OLI images collected from the same locations and same day as the Dove images. As a visual demonstration, Fig. 9 shows the spatial distributions of *Sargassum* abundance in each 1^o grid over the two days. To have an apples-to-apples comparison, the results shown in each grid are from the common areas where both sensors have valid measurements. The ratio between the two sensors (MODIS/Dove) on their estimated *Sargassum* abundance in each grid from their common areas is also shown in Fig. 9. From the ratio images, it is clear that, in most cases, MODIS estimates are lower than Dove estimates (i.e., ratio < 1.0), especially in the western GOM when the *Sargassum* amount is relatively low. In the eastern GOM where both sensors show higher *Sargassum* amounts than in the western GOM, MODIS estimates can occasionally exceed Dove estimates, likely due to mismatch between the two measurements over the fast moving *Sargassum* features under the influence of the Loop Current. Overall, from their common valid areas, on 5 June 2018 Dove detected ~54.7 km^2 of *Sargassum*, ~200% greater than the MODIS detection (~18.4 km^2). On 3 June 2019, Dove detected 50.0 km^2 of *Sargassum*, ~160% more than the MODIS detection (19.3 km^2).

The underestimation by MODIS can also be quantified statistically, as shown in Fig. 10a. From the 37 1^o grids, on average, Dove shows 156% higher *Sargassum* estimates than MODIS. Although the number varies across different grids, these comparisons clearly show that, on average, *Sargassum* estimates from MODIS images should only represent a lower bound, as the "missing" *Sargassum* can be > 150% of those estimated from MODIS.

Similar comparisons can also be obtained between Dove and MSI, and between Dove and OLI for the common valid areas (Figs. 10b & c). Dove consistently observed 368% more *Sargassum* than

MSI and 69% more *Sargassum* than OLI. The difference between Dove and OLI is lower than between Dove and MSI, suggesting that OLI can detect more *Sargassum* than MSI. Indeed, Fig. 8c and Fig. 10d both show that the matching *Sargassum* features are more "detectable" on the OLI FAI than on the MSI FAI. This is mostly attributed to the higher SNRs of the OLI NIR bands (Pahlevan et al., 2017), as sub-pixel detection limit (in %) is primarily a function of the sensor's SNRs (Hu et al., 2015; Qi and Hu, 2021). If summed up over all the matching image pairs, the total *Sargassum* coverage derived from Dove is 10.0 km^2, ~368% higher than the MSI estimates (2.1 km^2). Similarly, the total *Sargassum* coverage derived from the matching Dove images is 29.1 km^2, ~70% larger than the OLI estimates (17.2 km^2).

The MSI and OLI images were also compared with MODIS observations to evaluate the cross-sensor uncertainties in *Sargassum* estimates. Forty-five MSI images (tile: T20PNC) and fourteen OLI images collected in 2018 near the Lesser Antilles Islands were compared with the same-day MODIS measurements over their common valid areas. The total *Sargassum* biomass in the match-up areas from MODIS and MSI or OLI were summarized in Figs. 10e-f.

Overall, the relationship between MSI and MODIS is less clear ($R^2 = 0.61$, Fig. 10e) than between OLI and MODIS ($R^2 = 0.98$, Fig. 10f) or between MSI and OLI ($R^2 = 0.73$, Fig. 10d). The data spread in the MSI-MODIS relationship is similar to the observations in Wang and Hu (2020) where a different region and a different extraction method were used to extract *Sargassum* features from MSI images. As shown in Wang and Hu (2020), the potential reasons behind the data spread could be related to the finer MSI spatial resolution and the false-negative detection of small *Sargassum* features on MSI images. In contrast, the *Sargassum* estimates from OLI and MODIS are very consistent ($R^2 = 0.98$, Fig. 10f), although the biomass estimated from OLI is mostly higher than from MODIS. If summed up from the listed matching image pairs, OLI detected 62.3 kilotons of *Sargassum*, ~29% higher than the MODIS estimates (55.2 kilotons).

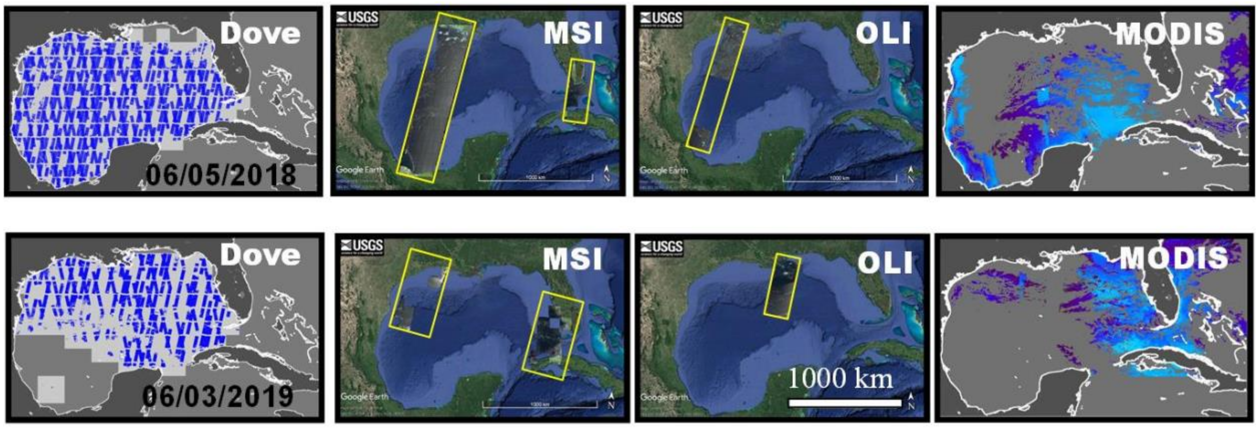
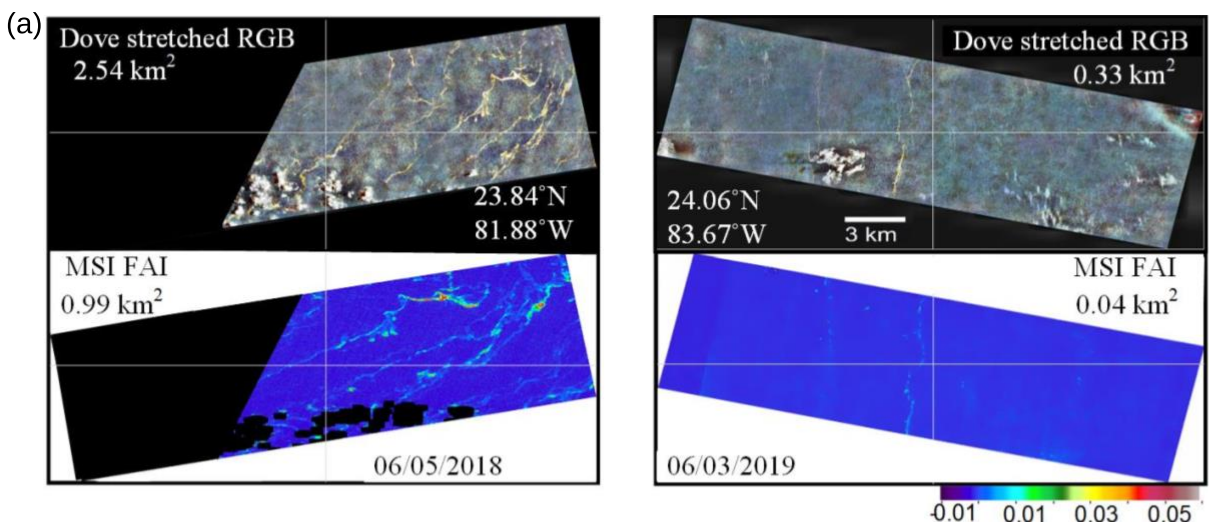


Fig. 7. Comparison of daily coverages of the data collected by Dove, MSI, OLI, and MODIS sensors on 5 June 2018 (top row) and on 3 June 2019 (bottom row). The purple to blue color and the yellow boxes highlight the areas where Dove, MSI, OLI, or MODIS has valid measurements.



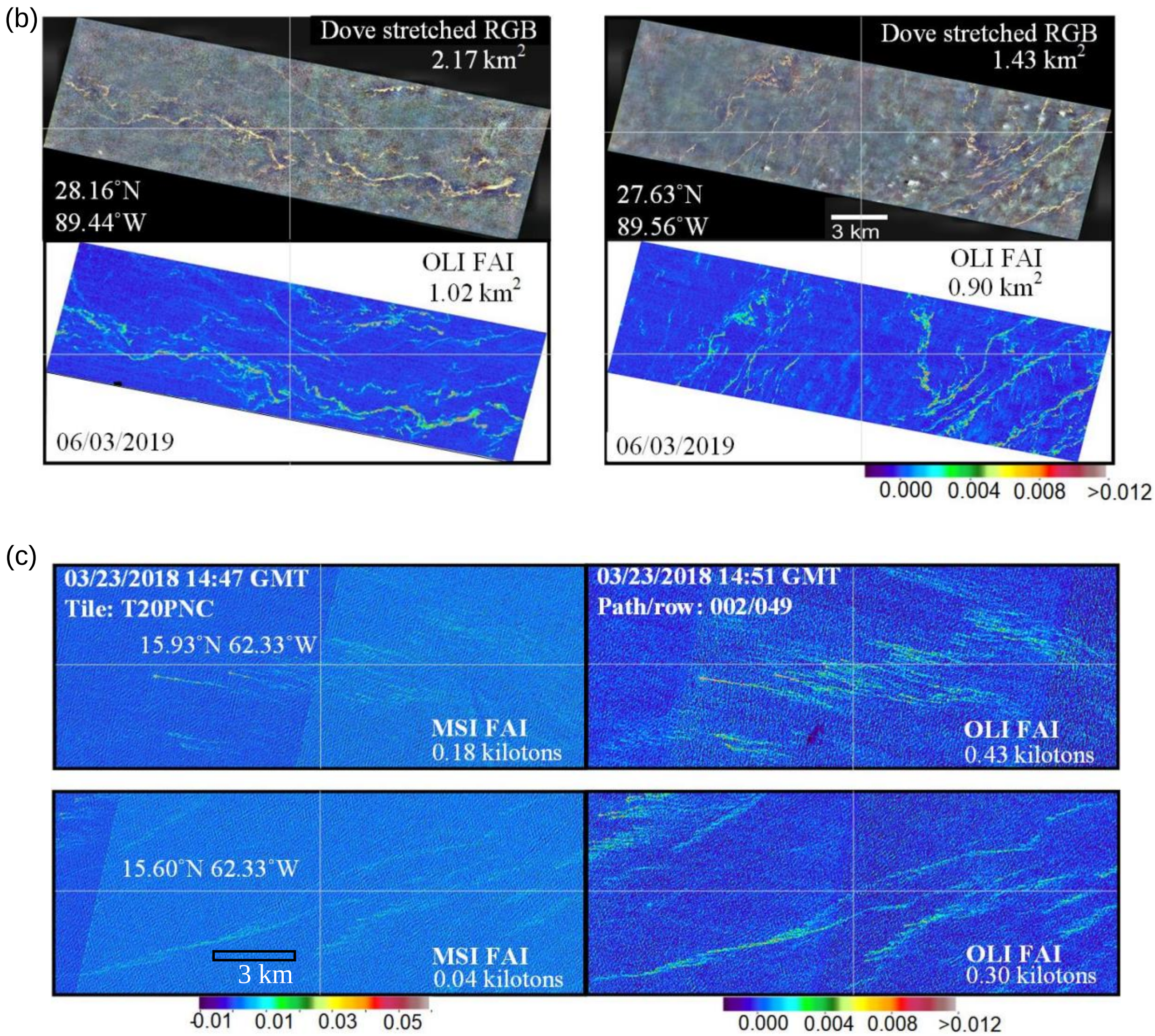


Fig. 8. (a-c) Comparison of image characteristics of quasi-simultaneous MSI, OLI and Dove image pairs, where the MSI and OLI images are cropped to match the same-day Dove images. The Dove images are color stretched using a contrast limited adaptive histogram equalization to enhance the contrast between *Sargassum* features and background water. The center coordinates and the extracted *Sargassum* biomass or areal coverages are labeled on the corresponding images. No *Sargassum* extraction method was applied to these images.

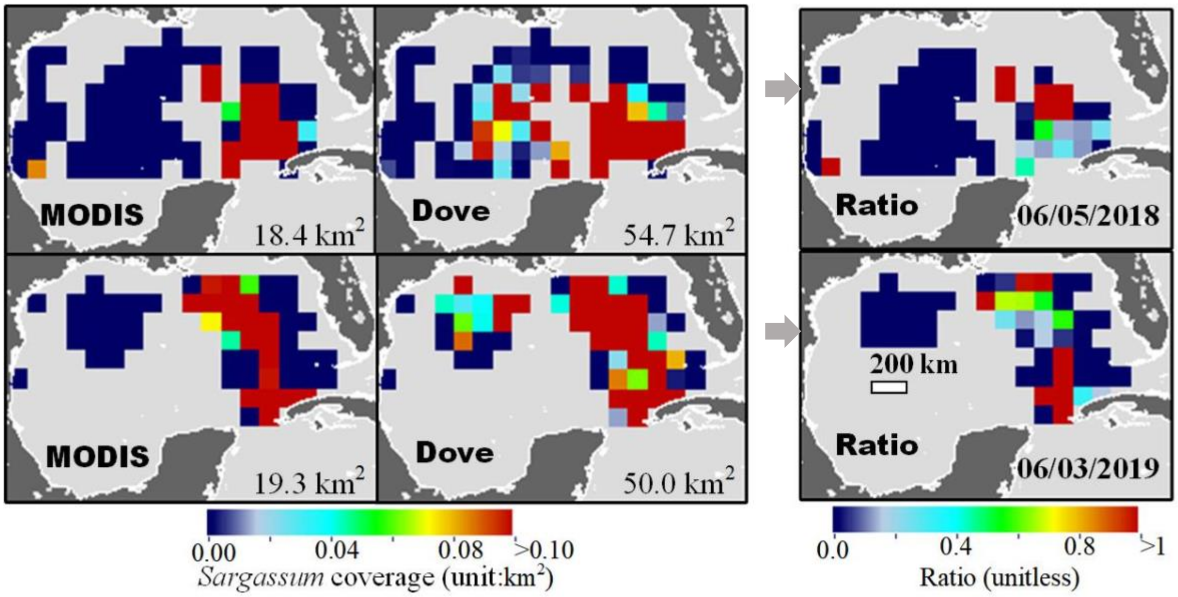


Fig. 9. Comparison of *Sargassum* coverage derived from MODIS and Dove over each 1˚ grid for 5 June 2018 and for 3 June 2019. The total *Sargassum* coverage measured from the common areas are marked on the bottom right corners of MODIS and Dove images. The ratio images show the ratios of MODIS/Dove derived *Sargassum* areas. Note that the data coverages of the ratio images are smaller than the common areas because on some of the common grids Dove measured no *Sargassum* (the denominator) and their ratio values are invalid. On 5 June 2018 and 3 June 2019, the common areas are ~150,000 km^2 and ~80,000 km^2, respectively. Gray color indicates non-common area due to no valid observations from either sensor. The *Sargassum* areal coverages in many grids are much higher than the top color of 0.1 km^2 (the highest value here is ~17 km^2). This colorbar was applied to emphasize the scattered small *Sargassum* features in the western GOM that are not observed in MODIS images. The match-up data are used to derive the relationships shown in Fig. 10a.

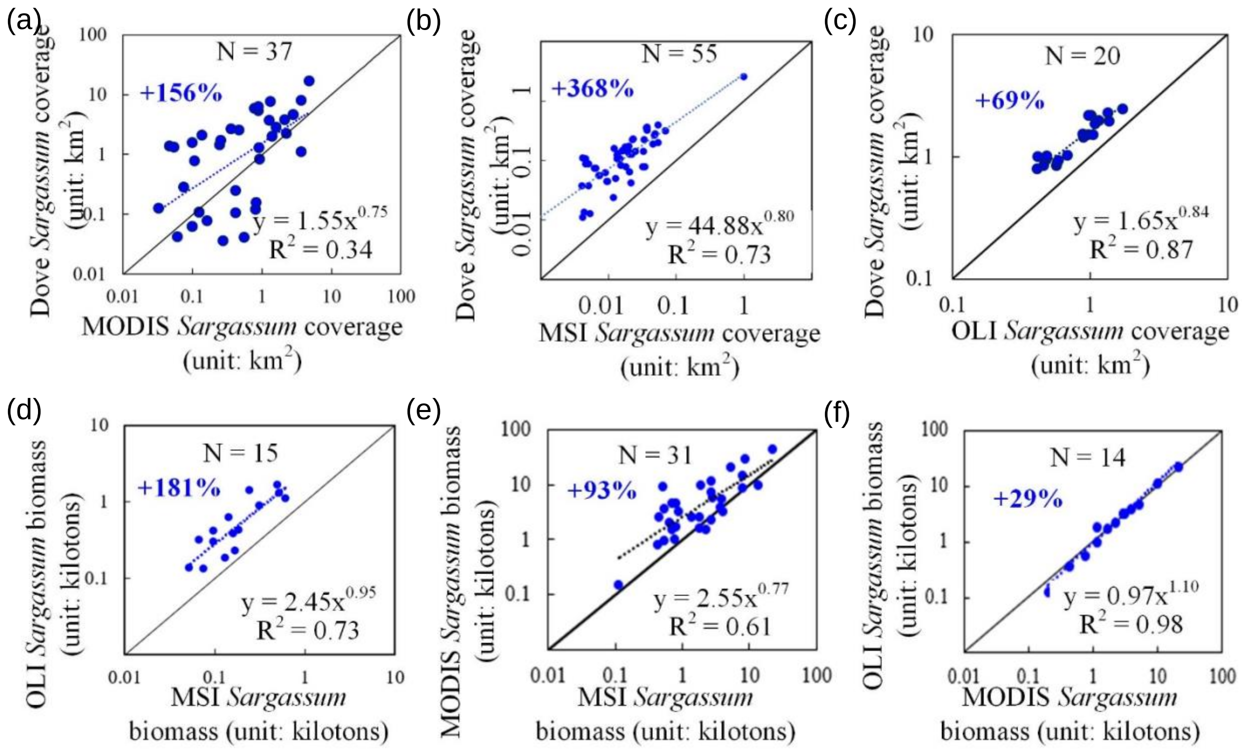


Fig. 10. Comparison of *Sargassum* biomass or coverages estimated from quasi-simultaneous MODIS, MSI, OLI, and Dove image pairs. Each dot represents the result from one pair of images. The black lines are the 1:1 lines, while the dotted blue lines are the linear fits in log space. The number of image pairs and the fitted equations are marked on the corresponding scatter plots. The relative differences (i.e., (y-x)/x) of the total *Sargassum* amount or area estimated from the two sensors in all the matching points are labeled in blue. Note that *Sargassum*-containing pixels from Dove are assumed to have 100% sub-pixel coverage while from other sensors are weighted by the sub-pixel coverage using the linear unmixing model.

3.3 **Size, biomass, and morphology of *Sargassum* features observed from MSI, OLI, and Dove images**

In addition to *Sargassum* abundance and distribution, characteristics of individual *Sargassum* features are also important for a number of reasons, for example to help implement plans for physical removal. Similar to Wang and Hu (2020), this study uses the following parameters to characterize individual features: biomass (kg), size (m^2), length (m), and length/width ratio. Fig. 11 shows that these parameters differ among the three sensors. Here, 22 MSI and 16 OLI images collected near the Lesser Antilles Islands (tile T20PNC and path/row: 001/050) and 4,375 Dove images collected in the GOM on 3 June 2019 were used to characterize *Sargassum* features. The feature morphology (size, length, length/width ratio) was calculated after applying a morphological close operation using a 3×3 pixel window. Then, for MSI and OLI, biomass of each feature was estimated using the corresponding FAI-biomass model (Wang et al., 2018; Wang & Hu, 2020). For Dove, biomass of each feature was estimated from the areal coverage after applying a conversion factor derived from the field (3.34 kg/m^2, Wang et al., 2018).

As shown in Fig. 11, the number of *Sargassum* features decreases sharply with increasing size and biomass. Although the size and length of average features from OLI are both much higher than from MSI, the average biomass per feature is rather similar between the two sensors, suggesting that that biomass density in the "extra" *Sargassum* area in OLI images (compared to MSI images) is rather low. This is because of the higher SNRs of OLI than MSI. Overall, with finer resolution, Dove-detected *Sargassum* features are much smaller, and their corresponding biomass per feature is also much lower. Because of the finer resolution, these characteristics are closer to the truth than those estimated from OLI or MSI. In contrast, regardless of the resolution, *Sargassum* are consistently observed as elongated features with mean length/width ratios of 3 – 5.

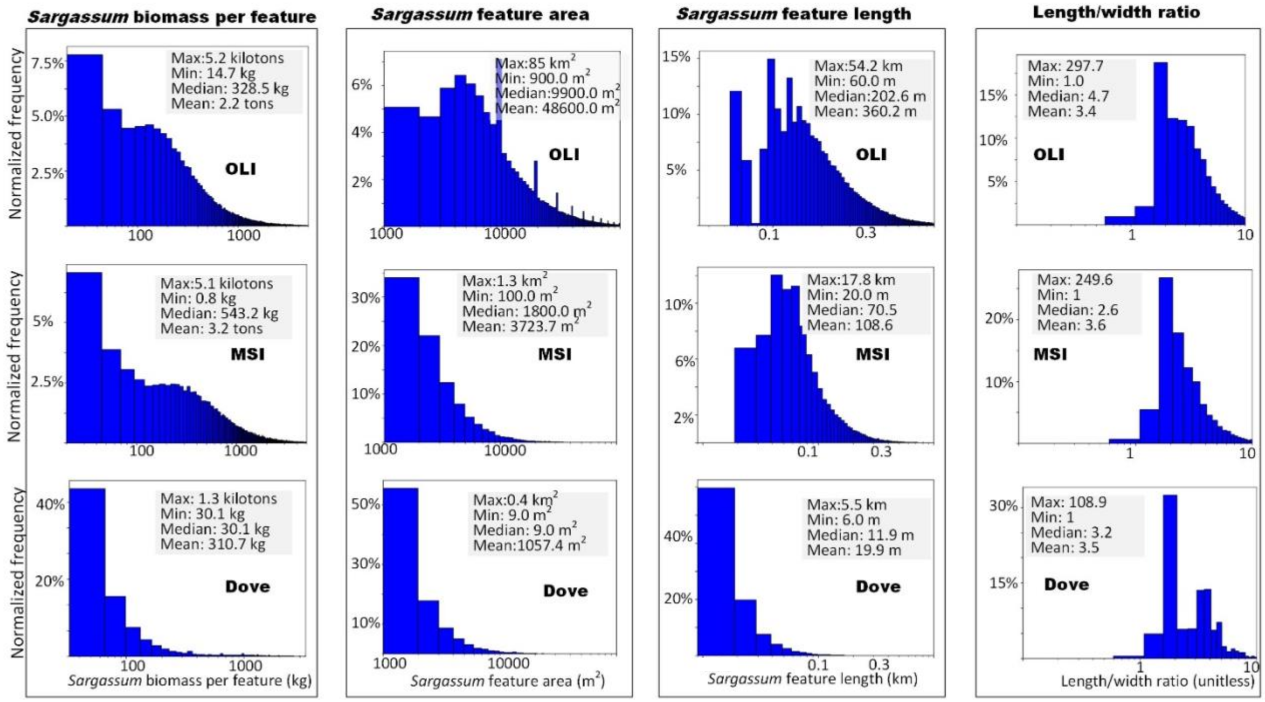


Fig. 11. Characteristics of individual *Sargassum* features derived from OLI (N = 16), MSI (N = 22), and Dove (N = 4,375) images. For each dataset, the normalized distributions of *Sargassum* biomass per feature, feature size, feature length, and length/width ratio are plotted. The maximum, minimum, median, and mean values are annotated on the corresponding plots. Most *Sargassum* features have relatively small size and low biomass, and smaller features are detected from the finer-resolution Dove images than from other sensors.

4. Discussions

4.1. Where is the "truth"

Because of the large spatial and temporal coverages, satellite remote sensing is perhaps the most reliable technique to observe large-scale *Sargassum* distributions and long-term changes. However, because many *Sargassum* clumps or rafts are small and moving in the ocean (Butler, Morris, Cadwallader, J., Stoner, 1983; Ody et al, 2019), it is nearly impossible to measure *Sargassum* size and biomass in the field to match satellite pixels, and therefore it is extremely difficult to validate satellite estimates in a quantitative way through field measurements. This is similar to mapping other small features such as oil slicks and *Ulva* macroalgae. Then, how can the uncertainties be quantified and how much *Sargassum* may be "missed" in the satellite observations? In other words, where is the "truth"?

Assuming that high-resolution sensors may provide estimates closer to the "truth", one way to quantify uncertainties in coarse-resolution estimates is through comparison of the two, as shown in several previous studies (Hu et al, 2016b; Cui et al. 2018; Hu et al., 2019; Wang et al., 2021). Yet these studies only used a few images for the comparison due to the difficulties in analyzing the high-resolution data in an automatic fashion, making it hard to draw conclusions. The PlanetScope constellation is the only data source at 3-m resolution with daily coverage of the entire GOM, thus providing an excellent opportunity to evaluate uncertainties in the *Sargassum* estimates from coarse-resolution sensors. Using 12,024 Dove images as the reference, it was determined that all MODIS, MSI, and OLI sensors underestimated *Sargassum* coverage and biomass. Overall, Dove showed ~150%, ~360%, and ~70% more *Sargassum* than MODIS, MSI, and OLI, respectively, when assuming *Sargassum* pixels detected on Dove RGB images all have 100% coverage. However, these numbers all depend on whether most *Sargassum* in the ocean comes from large, dense rafts or smaller clumps, and therefore would vary with time and location.

The same argument also applies to Dove images, as some small *Sargassum* features may still be undetected in the 3-m Dove images. For this reason, the Dove estimates are not the "truth" itself, but can only be regarded as being closer to the "truth". In fact, Dove estimates should only represent a lower bound of the true (actual) *Sargassum* abundance in the natural environment. In future studies, sensors with higher resolution or higher SNRs than Dove may be explored further to push the limit of satellite remote sensing of *Sargassum* and other macroalgae.

4.2. Advantages and disadvantages of the VGGUnet model for *Sargassum* extraction

This work is possible because of the use of deep learning techniques on the 3-band Dove images and other high-resolution images. Otherwise, due to the lack of spectral bands in the NIR wavelengths it is nearly impossible to extract accurate *Sargassum* features from the 12,025 images where confusion features such as clouds and cloud shadows are often found. Compared to the traditional machine-learning methods, deep learning techniques have the advantages of being a fast and reliable way to interpret vast amounts of satellite data (Figs. S2-S5 and Tables S2-S5). This is because many noisy features are spectrally similar to weak *Sargassum* features, and the traditional machine-learning methods only rely on the spectral information while the VGGUnet model relies also on the spatial context. Using the unique network structure, the VGGUnet model shows **robust performance** even with limited spectral bands, large background variations, and various confusing targets. This is especially important for high-resolution images where "noises" are highly variable, for example on the Dove images (see Fig S5 where the noises may be induced from clouds, sun glint, or wave glitter etc.). As the surface reflectance products are now provided from Planet Labs (Collison and Wilson, 2017), future studies may benefit from the improved data quality of Dove for better extracting *Sargassum* features using the deep learning methods. It is also noted that even when there are small errors in the manually prepared training data, the VGGUnet method can still be optimized to achieve satisfactory performance without bias. This is attributed to its ability to be trained to use not only the spectral information, but also the spatial context.

Another critical advantage of the proposed method is **its flexibility**. As demonstrated in this study, the VGGUnet model is easy to adapt to different type of satellite data or features through adjusting the input layer and optimizing the model parameters. The model input can be either single-band (e.g., FAI) or multispectral (e.g., RGB) images, depending on the specific feature characteristics. Likewise, when provided representative spatial and spectral patterns, the same model may also be trained to detect other image features such as clouds, whiting, oil slicks, and other macroalgal blooms including *Ulva prolifera* blooms, and features on different satellite images such as the coarse-resolution MODIS/VIIRS images (results not shown here). Because of its high flexibility, the method might also be applied to discriminate seagrass from *Sargassum* in coastal environment, especially when considering that seagrass features are more spatially diffuse than *Sargassum* slicks.

Moreover, this extraction model **requires no threshold** in detecting the *Sargassum* features. The decision is purely made with the optimized model weights learned from the training processes. This would reduce the potential biases due to the selection of extraction thresholds during the traditional threshold-based segmentation (Wang & Hu, 2016; Hu et al, 2019).

Finally, one disadvantage of the deep learning model is that the results can be difficult to interpret or diagnose, and the performance of the model strongly depends on the selection of representative training data. In contrast, for the traditional *Sargassum* detection methods, the rules for detecting the *Sargassum* features are more straightforward to understand, making it relatively easier to diagnose errors and improve performance (Wang & Hu, 2020). Nevertheless, with proper network structure and carefully selected training data, the deep learning model can greatly facilitate the use of vast amounts of high-resolution data in feature detection.

4.3. Near real-time *Sargassum* monitoring and tracking in nearshore waters

The availability of the various types of high-resolution data, combined with the success of the VGGUnet model in extracting *Sargassum* features automatically, makes it possible to fill the data gaps in nearshore waters from the coarse-resolution *Sargassum* imagery products (Fig. 1b). For example, corresponding to the MSI FAI image in the nearshore waters around Florida Keys (Fig.

1c), the extraction results in Fig. 12a clearly reveal *Sargassum* slicks with fine details. Besides that, Fig. 12b shows an example of the *Sargassum* slicks extracted from the 3-m Dove images collected in the same area with even finer details. While the latency between satellite overpass and data access is often less than a day, whether or not a near real-time system can be established to fully use the high-resolution data depends on the processing speed, as high-resolution data have much higher data volume (e.g., for the same area, a 3-m Dove image has >110,000 times more pixels than the corresponding 1-km MODIS image).

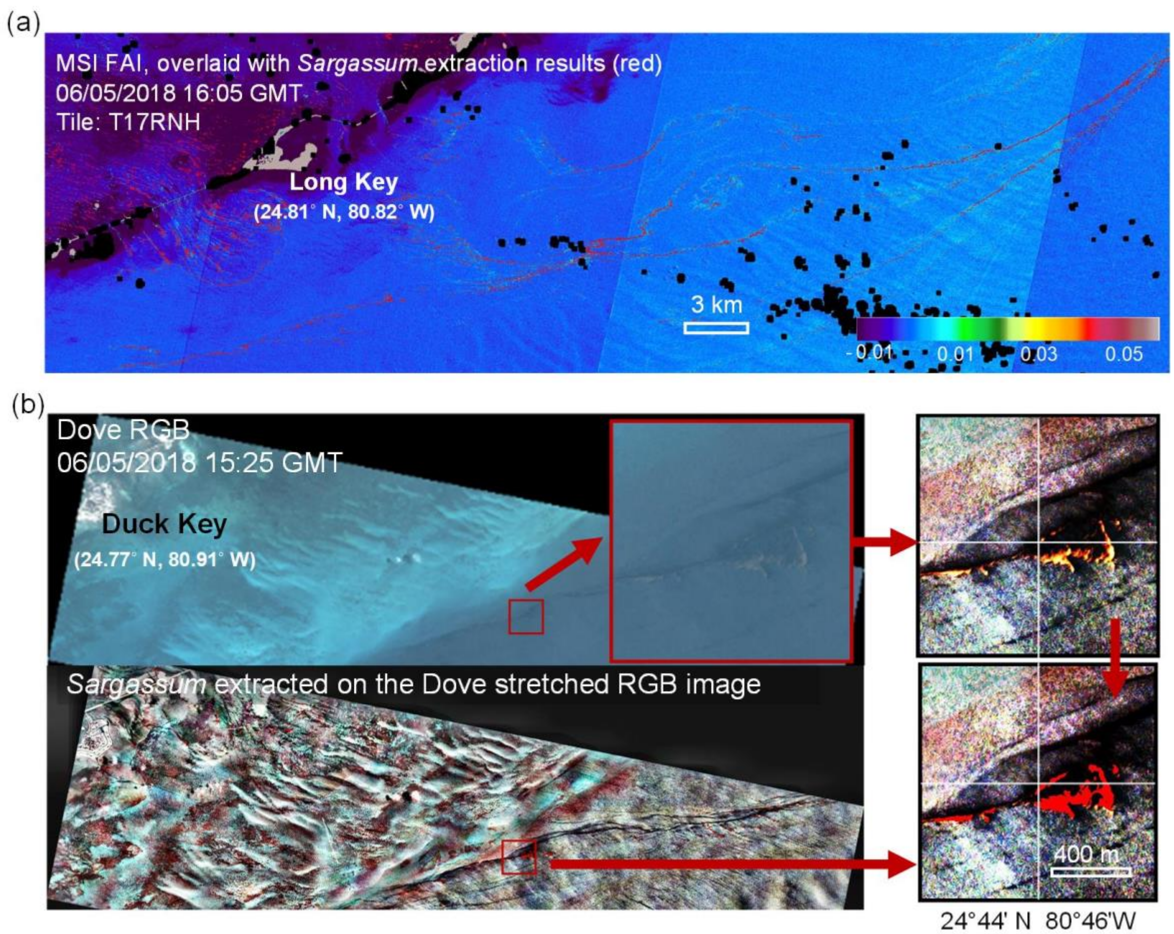


Fig. 12. *Sargassum* features extracted from high-resolution MSI and Dove images near the coast of Florida Keys. (a) MSI FAI image near Long Key in the Florida Keys, with *Sargassum* extraction results overlaid in red. The color legend applies to FAI values. A portion of this image is shown in Fig. 1c. (b) Dove RGB and stretched RGB images on the same day of the MSI image near Duck Key in the Florida Keys. The sub-images to the right are the Dove stretched RGB images enlarged from the red box, where

Sargassum extraction results are overlaid in red. The central coordinates of the sub-image are labeled below the image.

Table 4 summarizes the approximate processing speed for *Sargassum* extraction from individual MSI, OLI, and Dove images. For an MSI FAI image with 10,000 × 10,000 pixels, the *Sargassum* extraction time using the VGGUnet model is about 2 minutes (123.0 seconds), much lower than the time needed by the previous method where the TNRD denoising process alone takes about 11 minutes (651 seconds, Wang & Hu, 2020). For OLI and Dove images, because the image sizes in terms of number of pixels are slightly smaller than MSI images, they require less time to extract the *Sargassum* features using the VGGUnet model (Table. 4). For a coastal region of $1^{o} \times 1^{o}$ in the tropical or subtropical ocean, it takes about 42 Dove images and 71 minutes to process all images, thus meeting the requirement of near real-time monitoring. For the same $1^{o} \times 1^{o}$ region, it takes only 2 minutes and 22 seconds to process one MSI and one OLI image, respectively.

A near real-time monitoring system also requires frequent **data coverage**. While MSI and OLI show better *Sargassum* extraction accuracy than Dove when their own images are used as the reference, only the latter can provide daily coverage. The 3-m resolution also makes it possible to "see" cloud-free pixels among small clouds, thus improving the spatial coverage. Therefore, a combination of all available Dove, MSI, and OLI images should be able to meet the critical requirement of a near real-time *Sargassum* monitoring and tracking system for targeted nearshore waters. We expect to implement such a capacity into the existing SaWS with the VGGUnet model in the near future.

Table 4. Approximate processing time for *Sargassum* extraction from MSI, OLI, and Dove images on a PC with IntelI CoreI i9-9900 CPU @ 3.30GHz and a Nvidia GeForce RTX 2080 Ti GPU. The estimated processing time is averaged over 22 MSI images with 10,686 × 10,866 pixels per image, 10 OLI images with 7,138 × 7,391 pixels per image, and 29 Dove images with ~8,000 × 4,000 pixels per image.

Sensor	**MSI**	**OLI**	**Dove**

Mean processing time per image	123.0 seconds	85.5 seconds	101.6 seconds

5. Conclusion

Using a deep convolutional neural network, this study designed a VGGUnet-based approach to automatically detect and quantify *Sargassum* macroalgae from various high-resolution images. Even with the complex ocean background and variable "noises", experiments on the MSI, OLI, WV-2, and Dove images all achieved high self-consistency detection accuracy with fast processing speeds. Overall, this work provides a generic (i.e., applicable to other features such as oil slicks), concise, and effective tool for extracting *Sargassum* features from high-resolution satellite images, and will also satisfy the needs for near real-time *Sargassum* bloom monitoring in coastal regions. The work also enables a first-ever systematic, statistically meaningful way to evaluate how much *Sargassum* is "missed" by coarse-resolution sensors such as MODIS. Depending on the locations, all MSI, OLI, and MODIS sensors may miss considerable amount of *Sargassum* as compared with concurrent and co-located Dove (3-m resolution) estimates. However, as long as the underestimates are systematic rather than random, previous long-term MODIS estimates should be valid for long-term trend studies. Finally, following this work, the high-resolution MSI, OLI, and Dove images, once incorporated into the existing *Sargassum* Watch System, are expected to make significant improvements by filling the important data gaps in nearshore waters on a daily basis.

Acknowledgement:

This work was supported by the U.S. NASA Ocean Biology and Biogeochemistry program NNX16AR74G, 80NSSC20M0264), Ecological Forecast program (NNX17AF57G, 80NSSC21K0422), Water Resources program (80NSSC19K1200), NOAA RESTORE Science

Program (NA17NOS4510099), and the JPSS/NOAA Cal/Val project (NA15OAR4320064). We thank NASA, USGS, Digital Globe, and Planet Labs, Inc. providing satellite data. DigitalGlobe and Planet Dove data were provided by the Commercial Archive Data for NASA investigators (cad4nasa.gsfc.nasa.gov) under the National Geospatial-Intelligence Agency's NextView license agreement. We thank Shaojie Sun (USF) for downloading all Dove data used in this study. We thank Brock Murch (USF) for copy editing the manuscript, and thank the three anonymous reviewers for providing valuable comments.

References:

Arellano-Verdejo, J., Lazcano-Hernandez, H. E., & Cabanillas-Terán, N. (2019). ERISNet: deep neural network for *Sargassum* detection along the coastline of the Mexican Caribbean. *PeerJ*, *7*, e6842.

Bach, L. T., Tamsitt, V., Gower, J., Hurd, C. L., Raven, J. A., & Boyd, P. W. (2021). Testing the climate intervention potential of ocean afforestation using the Great Atlantic Sargassum Belt. *Nature communications*, *12*(1), 1-10.

Butler, J. N., Morris, B. F., Cadwallader, J., Stoner, A. W. (1983) Studies of *Sargassum* and the *Sargassum* community. *Bermuda Biological Station for Research*, 22.

Chávez, V., Uribe-Martínez, A., Cuevas, E., Rodríguez-Martínez, R. E., van Tussenbroek, B. I., Francisco, V., ... & Silva, R. (2020). Massive Influx of Pelagic *Sargassum* spp. on the Coasts of the Mexican Caribbean 2014–2020: Challenges and Opportunities. *Water*, *12*(10), 2908.

Chinchor, N., Sundheim, B. (1993, August). MUC-5: information extraction system evaluation. In *Proceedings of the 5th conference on Message understanding* (pp. 27-44). Association for Computational Linguistics.

Collison, A., & Wilson, N. (2017). Planet Surface Reflectance Product. *Technical White Paper: Version*, *1*.

Cui, T. W., Liang, X. J., Gong, J. L., Tong, C., Xiao, Y. F., Liu, R. J., ... & Zhang, J. (2018). Assessing and refining the satellite-derived massive green macro-algal coverage in the Yellow Sea with high resolution images. *ISPRS journal of photogrammetry and remote sensing*, *144*, 315-324.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). IEEE.

Diakogiannis, F. I., Waldner, F., Caccetta, P., & Wu, C. (2019). ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data. *arXiv preprint arXiv:1904.00592*.

Gower, J., & King, S. (2019a). Seaweed, seaweed everywhere. *Science*, *365*(6448), 27-27.

Gower, J., & King, S. (2019b). The distribution of pelagic *Sargassum* observed with OLCI. *International Journal of Remote Sensing*, 1-11.

Gower, J., Young, E., King, S. (2013). Satellite images suggest a new *Sargassum* source region in 2011. Remote Sensing Letters, 4, 764–773.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Hedley, J. D., Roelfsema, C., Brando, V., Giardino, C., Kutser, T., Phinn, S., ... & Koetz, B. (2018). Coral reef applications of Sentinel-2: Coverage, characteristics, bathymetry and benthic mapping with comparison to Landsat 8. *Remote sensing of environment*, *216*, 598-614.

Hordiiuk, D., Oliinyk, I., Hnatushenko, V., & Maksymov, K. (2019, April). Semantic Segmentation for Ships Detection from Satellite Imagery. In *2019 IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO)* (pp. 454-457). IEEE.

Hu, C., Barnes, B. B., Murch, B., & Carlson, P. R. (2013). Satellite-based virtual buoy system to monitor coastal water quality. *Optical Engineering*, *53*(5), 051402.

Hu, C., Feng, L., Hardy, R. F., & Hochberg, E. J. (2015). Spectral and spatial requirements of remote measurements of pelagic *Sargassum* macroalgae. *Remote Sensing of Environment*, *167*, 229-246.

Hu, C., Murch, B., Barnes, B. B., Wang, M., Maréchal, J. P., Franks, J., ... & Siuda, A. (2016a). *Sargassum* watch warns of incoming seaweed. *Eos*, *97*, 10-15.

Hu, C., Hardy, R., Ruder, E., Geggel, A., Feng, L., Powers, S., ... & McDonald, T. (2016b). *Sargassum* coverage in the northeastern Gulf of Mexico during 2010 from Landsat and airborne observations: Implications for the Deepwater Horizon oil spill impact assessment. *Marine pollution bulletin*, *107*(1), 15-21. doi: 10.1016/j.marpolbul.2016.04.045.

Hu, L., Zeng, K., Hu, C., & He, M. X. (2019). On the remote estimation of *Ulva prolifera* areal coverage and biomass. *Remote sensing of environment*, *223*, 194-207.

Iglovikov, V., Mushinskiy, S., & Osin, V. (2017). Satellite imagery feature detection using deep convolutional neural network: A kaggle competition. *arXiv preprint arXiv:1706.06169*.

Iglovikov, V., & Shvets, A. (2018). Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. *arXiv preprint arXiv:1801.05746*.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Langin K. (2018). Mysterious masses of seaweed assault Caribbean islands. doi:10.1126/science.aau4441. https://www.sciencemag.org/news/2018/06/mysterious-masses-seaweed-assault-caribbean-islands.

Li, X., Liu, B., Zheng, G., Ren, Y., Zhang, S., Liu, Y., ... & Wang, F. (2020). Deep learning-based information mining from ocean remote sensing imagery. *National Science Review*.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, 152, 166-177.

Ody, A., Thibaut, T., Berline, L., Changeux, T., André, J. M., Chevalier, C., ... & Connan, S. (2019). From *In Situ* to satellite observations of pelagic *Sargassum* distribution and aggregation in the Tropical North Atlantic Ocean. *PloS one*, *14*(9).

Pahlevan, N., Sarkar, S., Franz, B. A., Balasubramanian, S. V., & He, J. (2017). Sentinel-2 MultiSpectral Instrument (MSI) data processing for aquatic science applications: Demonstrations and validations. *Remote sensing of environment*, *201*, 47-56.

Qi, L., & Hu, C. (2021). To what extent can *Ulva* and *Sargassum* be detected and separated in satellite imagery? *Harmful Algae*, 103, 102001, https://doi.org/10.1016/j.hal.2021.102001

Rodríguez-Martínez, R. E., Medina-Valmaseda, A. E., Blanchon, P., Monroy-Velázquez, L. V., Almazán-Becerril, A., Delgado-Pech, B., ... García-Rivas, M. C. (2019). Faunal mortality associated with massive beaching and decomposition of pelagic *Sargassum*. *Marine Pollution Bulletin*, *146*, 201-205.

Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Siuda, A. N., Schell, J. M., Goodwin, D. S. (2016). Unprecedented proliferation of novel pelagic *Sargassum* form has implications for ecosystem function and regional diversity in the Caribbean. In *American Geophysical Union, Ocean Sciences Meeting 2016, abstract# ME14E-0682*.

Smetacek, V., & Zingone, A. (2013). Green and golden seaweed tides on the rise. *Nature*, *504*(7478), 84-88.

van Tussenbroek, B. I., Arana, H. A. H., Rodríguez-Martínez, R. E., Espinoza-Avalos, J., Canizales-Flores, H. M., González-Godoy, C. E., ... Collado-Vides, L. (2017). Severe impacts of brown tides caused by *Sargassum* spp. on near-shore Caribbean seagrass communities. *Marine pollution bulletin*, *122*(1-2), 272-281.

Wang, M., & Hu, C. (2015). Extracting oil slick features from VIIRS nighttime imagery using a Gaussian filter and morphological constraints. *IEEE Geoscience and Remote Sensing Letters*, *12*(10), 2051-2055.

Wang, M., & Hu, C. (2016). Mapping and quantifying *Sargassum* distribution and coverage in the Central West Atlantic using MODIS observations. *Remote sensing of environment*, *183*, 350-367.

Wang, M., Hu, C., Barnes, B. B., Mitchum, G., Lapointe, B., & Montoya, J. P. (2019). The great Atlantic *Sargassum* belt. *Science*, *365*(6448), 83-87.

Wang, M., Hu, C., Cannizzaro, J., English, D., Han, X., Naar, D., ... & Hernandez, F. (2018). Remote sensing of *Sargassum* biomass, nutrients, and pigments. *Geophysical Research Letters*, *45*(22), 12-359.

Wang, M., & Hu, C. (2020). Automatic extraction of *Sargassum* features from Sentinel-2 MSI Images. *IEEE Transactions on Geoscience and Remote Sensing*, doi: 10.1109/TGRS.2020.3002929.

Wang, S., Liu, L., Qu, L., Yu, C., Sun, Y., Gao, F., & Dong, J. (2019). Accurate *Ulva prolifera* regions extraction of UAV images with superpixel and CNNs for ocean environment monitoring. *Neurocomputing*, *348*, 158-168.

Wang, X., Xing, Q., An, D., Meng, L., Zheng, X., Jiang, B., & Liu, H. (2021). Effects of Spatial Resolution on the Satellite Observation of Floating Macroalgae Blooms. *Water*, *13*(13), 1761.

Webster, R. K., & Linton, T. (2013). Development and implementation of *Sargassum* early advisory system (SEAS). Shore & Beach, 81(3), 1.

Wicaksono, P., & Lazuardi, W. (2018). Assessment of PlanetScope images for benthic habitat and seagrass species mapping in a complex optically shallow water environment. *International journal of remote sensing*, *39*(17), 5739-5765.

Xiao, X., Lian, S., Luo, Z., & Li, S. (2018, October). Weighted Res-UNet for High-Quality Retina Vessel Segmentation. In *2018 9th International Conference on Information Technology in Medicine and Education (ITME)* (pp. 327-331). IEEE.

Xing, Q., Wu, L., Tian, L., Cui, T., Li, L., Kong, F., ... & Wu, M. (2018). Remote sensing of early-stage green tide in the Yellow Sea for floating-macroalgae collecting campaign. *Marine pollution bulletin*, *133*, 150-156.

Xing, Q., An, D., Zheng, X., Wei, Z., Wang, X., Li, L., ... & Chen, J. (2019). Monitoring seaweed aquaculture in the Yellow Sea with multiple sensors for managing the disaster of macroalgal blooms. *Remote Sensing of Environment*, *231*, 111279.

Zheng, Y., Wu, J., Wang, A., & Chen, J. (2018). Object-and pixel-based classifications of macroalgae farming area with high spatial resolution imagery. *Geocarto International*, *33*(10), 1048-1063.