doi:10.20944/preprints202103.0780.v1

Article

# Semi-supervised segmentation for coastal monitoring seagrass using RPA imagery

**Brandon Hobley [1,*], Riccardo Arosio [2], Geoffrey French[1], Julie Bremner [2], Tony Dolphin [2] and Michal Mackiewicz [1]**

[1] School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, UK; g.french@uea.ac.uk (G.F.); m.mackiewicz@uea.ac.uk (M.M.)
[2] Collaborative Centre for Sustainable Use of the Seas. School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ; tony.dolphin@cefas.co.uk (T.D.); julie.brenmer@cefas.co.uk (J.B.); riccardo.arosio@cefas.co.uk (R.A.)

**Abstract:** Intertidal seagrass plays a vital role in estimating the overall health and dynamics of coastal environments due to its interaction with tidal changes. However, most seagrass habitats around the globe have been in steady decline due to human impacts, disturbing the already delicate balance in environmental conditions that sustain seagrass. Miniaturization of multi-spectral sensors has facilitated very high resolution mapping of seagrass meadows, which significantly improve the potential for ecologists to monitor changes. In this study, two analytical approaches used for classifying intertidal seagrass habitats are compared: Object-based Image Analysis (OBIA) and Fully Convolutional Neural Networks (FCNNs). Both methods produce pixel-wise classifications in order to create segmented maps, however FCNNs are an emerging set of algorithms within Deep Learning with sparse application towards seagrass mapping. Conversely, OBIA has been a prominent solution within this field, with many studies leveraging in-situ data and multiscale segmentation to create habitat maps. This work demonstrates the utility of FCNNs in a semi-supervised setting to map seagrass and other coastal features from an optical drone survey conducted at Budle Bay, Northumberland, England. Semi-supervision is also an emerging field within Deep Learning that has practical benefits of achieving state of the art results using only subsets of labelled data. This is especially beneficial for remote sensing applications where in-situ data is an expensive commodity. For our results, we show that FCNNs have comparable performance with standard OBIA method used by ecologists, while also noting an increase in performance for mapping ecological features that are sparsely labelled across the study site.

**Keywords:** Deep learning; Computer vision; Remote sensing; Supervised learning; Semi-supervised learning; Segmentation; Seagrass mapping

## 1. Introduction

Accurate and efficient mapping of seagrass extents is a critical task given the importance of these ecosystems in coastal settings and their use as a metric for ecosystem health. In particular, seagrass ecosystems play a key role for estimating and assessing the health and dynamics of coastal ecosystems due to their sensitive response to tidal processes ([1], [2], [3], [4]); or human-made artificial interference ([5], [6], [7]). Furthermore, seagrass plays a vital part in sediment stabilization [8], pathogen reduction [9], carbon sequestration ([10], [11]) and as a general indicator for water quality [12]. However, there is evidence seagrass areas have been in steady decline due to human disturbance for decades [13].

In coastal monitoring, Remote Sensing has provided a major platform for ecologists to assess and monitor sites for a plethora of applications [14]. Traditionally, passive remote sensing via satellite imagery was used to provide global to regional observations at regular sampling intervals, however it often struggles to overcome problems such as cloud contamination, oblique views and costs for data acquisition [15]. Another problem with satellite imagery is its coarse resolution. The shift to remotely piloted aircraft (RPAs), or drones, and commercially available cameras, resolves resolution by collecting several overlapping very high resolution (VHR) images [16] and stitching the sensor outputs together using Structure from Motion techniques to create orthomosaics [17]. The benefits for using these instruments are two-fold: firstly, the resolution of imagery can be user controlled with respects to drone altitude; secondly, sampling intervals are more accessible when compared with data acquisitions from satellite imagery. Advances in passive remote sensing have allowed coastal monitoring of seagrass, intertidal macro-algae and other species in study sites such as: Pembrokeshire, Wales [16]; Bay of Mont St Michel, France [18]; a combined study of Giglio island and the coast of Lazio, Italy [19] and Kilkieran Bay, Ireland [20] with the latter using a hyperspectral camera. The main goal of these studies is to create a habitat map by classifying multi-spectral drone data into sets of meaningful classes such that the spatial distribution of ecological features can be assessed [21]. This work also aims to produce a habitat map of Budle Bay; a large (2 km$^2$) square estuary on the North Sea in Northumberland, England (55.625°N, 1.745°W). Two species of seagrass, namely *Zostera noltii* and *Angustifolia*, are of interest, however this work will also consider all other coastal features of algae and sediment recorded in an in-situ survey conducted by the Centre for Environment, Fisheries and Aquaculture Science (CEFAS) and the Environment Agency.

Object Based Image Analysis (OBIA) [22] is an approach for habitat mapping that starts by performing an initial segmentation that clusters pixels into image-objects by maximising heterogeneity between said objects and homogeneity within them. The image segmentation method used in OBIA is multiscale segmentation (MSS), a non-supervised region-growing segmentation method that provides the grounds for extracting textural, spatial and spectral features that can be used for subsequent semantic modelling ([23], [24]). For habitat mapping in coastal environments, OBIA has found successful applications by using auxiliary in-situ data, i.e. ground truth data via site visit, for semantic modelling ([25], [26], [27], [28], [29], [20], [16]). A standard approach is to overlay in-situ data on generated image-objects through MSS so that selected objects are used to extract features that can create Machine Learning models. Tuned models are then used to classify the remaining image-objects, thus creating a habitat map.

Developments in Computer Vision through the introduction of Fully Convolutional Neural Networks (FCNNs) ([30], [31]) can provide an alternative approach to OBIA. FCNNs have been found to beat state of the art results on a number of established datasets for semantic segmentation[1] ([32], [33]) through the effective use of Graphical Processing Units (GPUs) and data-driven approaches. However, the mentioned datasets and challenges often suit these network architectures due to the abundant amount of labelled data these provide. In coastal monitoring this is a costly endeavour possible through in-situ surveying, and therefore this work will also investigate the use of semi-supervision as an alternative approach to modelling without the need of a fully labelled data set.

Whilst the main goal is to produce a reliable habitat map detailing areas of seagrass, algae and sediment at Budle Bay, this work also assesses the difference in performance between FCNNs and OBIA. In order to assess both methods, each algorithm uses the same imagery and knowledge gathered from the mentioned in-situ survey. Furthermore, the use of semi-supervised techniques is investigated due to the nature of the dataset used for modelling FCNNs (section 2.4).

Section 2.4 will detail the data collection and pre-process necessary for FCNNs, both methods will be explained and tailored for the study site in sections 2.5 and 2.6, results are presented in section 3 and an analysis of these results is in section 4. We will answer the following research questions:

---

[1] An equivalent output to habitat mapping.

- Can FCNNs model high resolution aerial imagery from a small set of geographically referenced image shapes?
- How does performance compare with standard OBIA/GIS frameworks?
- How accurate is modeling *Zostera noltii* and *Angustifolia* along with all other relevant coastal features within the study site?

## 2. Methods

### *2.1. Study site*

The research was focused on Budle Bay, Northumberland, England (55.625°N, 1.745°W). The coastal site has one tidal inlet, with previous maps also detailing the same inlet ([34], [35], [36]). Sinuous and dendritic tidal channels are present within the bay, and bordering the channels are areas of seagrass and various species of macro-algae.

### *2.2. Data collection*

Figure 1 displays very high resolution orthomosaics of Budle Bay created using Agisoft's MetaShape [37] and structure from motion (SfM). SfM techniques rely on estimating intrinsic and extrinsic camera parameters from overlapping imagery [38]. A combination of appropriate flight planning in terms of altitude and aircraft speed, and the camera's field of view were important for producing good quality orthomosaics. Two sensors were used: a SONY ILCE-6000 camera with 3 wide banded filters for Red, Green and Blue channels and a ground sampling distance of approximately 3 cm (Figure 1, bottom right). And a MicaSense RedEdge3 camera with 5 narrow banded filters for Red (655-680 nms), Green (540-580 nms), Blue (459-490 nms), Red Edge (705-730 nms) and Near Infra-red (800-880 nms) channels and a ground sampling distance of approximately 8 cm (Figure 1, top right).

Each orthomosaic was orthorectified using respective GPS logs of camera positions and ground markers that were spread out across the site. This process ensures that both mosaics were well aligned with respect to each other, and also with ecological features present within the coastal site.

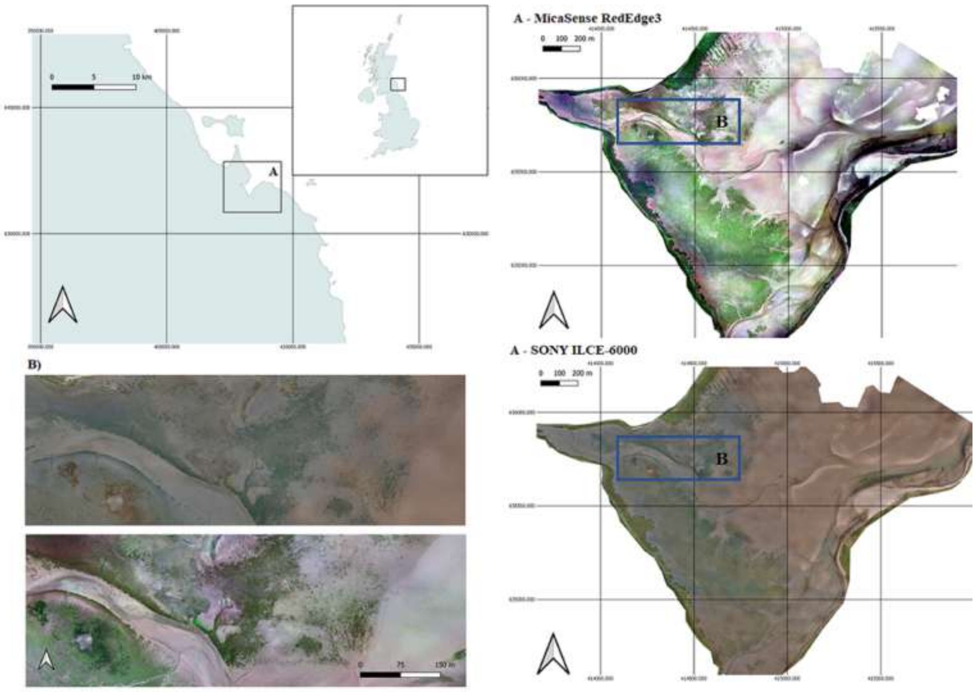

**Figure 1.** Study site within the U.K. (top-left). Ortho-registered images of Budle Bay using a SONY ILCE-6000 and a MicaSense RedEdge 3 (right). For the display of the latter camera, we use the Red, Green and Blue image bands.

### 2.3. On-situ survey

CEFAS and the Environment Agency conducted ground and aerial surveys of Budle Bay in September 2017 and noted 13 ecological targets that can be grouped into background sediment, algae, seagrass and saltmarsh. Classes within the background sediment were: rock, gravel, mud and wet sand. These features were modelled as one single class and dry sand was added to further aid distinguishing sediment features. Algal classes included *Microphytobenthos*, *Enteromorpha spp.* and other macroalgae (inc. *Fucus*). Lastly, the remaining coastal vegetation classes were seagrass and saltmarsh. Since the aim is to map the areas of seagrass in general, both species *Zostera noltii* and *Angustifolia* were merged to a single class while saltmarsh remains as a single class although two different species were noted. Thus, a total of 7 target classes can be listed.

- Background sediment: dry sand and other bareground
- Algae: *Microphytobenthos*, *Enteromorpha* and other macroalgae (including *Fucus*)
- Seagrass: *Zostera noltii* and *Angustifolia* merged to a single class
- Other plants: Saltmarsh

The in-situ survey recorded 108 geographically referenced tags with the percentage cover of all ecological features previously listed within a 300mm radius. These were dispersed mainly on the Western, Central and Southern portions of the site. Figure 2 displays the spatial distribution of recorded measurements by placing a point for each tag over the orthophoto generated using the SONY camera.

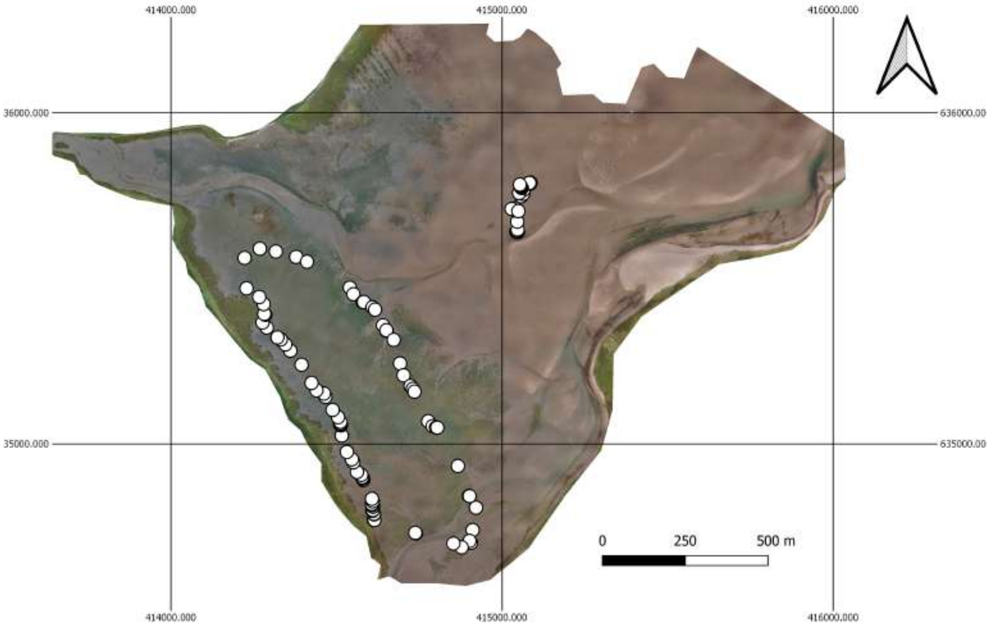

**Figure 2.** Distribution of recorded tags during the on-situ survey.

### 2.4. Data pre-process for FCNNs

The orthomosaic from the SONY camera was 87730×72328 pixels with 3 image bands orthomosaic, while the RedEdge3 multispectral orthomosaic was 32647×26534 with 5 image bands. For ease of processing, each orthomosaic was split into non-overlapping blocks of 6000×6000 images with each image containing geographic information to be used for further processing. The SONY orthomosaic was split into 140 tiles and the RedEdge3 into 24.

The recorded percentage covers were used to classify each point in Figure 2 to a single ecological class listed in section 2.3 based on the highest estimated cover during the in-situ survey. The classification for each point provides the basis to create geographically referenced polygon files through photo interpretation. The reasoning for using photo interpretation instead of selecting segmented image-objects was to avoid bias from the OBIA when generating segmentation maps used for FCNN training. The assessment of both algorithms was done using a subset of polygons to compare with predicted habitat maps from both methods. Figure 3 displays a gallery of images for each class with some example polygons.

### 2.4.1. Polygons to segmentation masks for FCNNs

Each polygon contains a unique semantic value depending on the recorded class. FCNNs were trained with segmentation maps that contain a one-to-one mapping of pixels encoded with a semantic value, with the goal to optimise this mapping [30]. Segmentation maps used for training FCNNs were created using the geographic coordinates stored in each polygon and converting real-world coordinates for each vertice to image-coordinates. If a polygon fits within an image, then the candidate image was sampled into 256×256 image tiles centered on labelled sections of the image. By cropping images centered on polygons the edges of each image have a number of pixels that were not labelled (Figure 5). The difference in spatial resolution for each camera results in a difference in labelled pixels, since each polygon covers the same area within the real-world. This process generated 534 images with the RedEdge3 multispectral camera which were split into 363 images for training, 102 images for testing and 69 images for validating and hyper-parameter tuning. The SONY camera produced 1108 images, split into 770 images for training, 213 images for testing and 125 for validation. The split of images for each camera is based off the associated split of polygons used for training/testing each method.

### 2.4.2. Vegetation, soil and atmospheric indices for FCNNs

Vegetation, soil and atmospheric indices are derivations from standard Red, Green and Blue and/or Near-infrared image bands that can aid discerning multiple vegetation classes [39]. Near-infrared and red, green and blue bands from the RedEdge3 were used to compute a variety of indices, adding 5 bands of data to each image. These extra bands were: Normalised Difference Vegetation Index (NDVI) [40], Atmospheric Resistant Vegetation Index (IAVI) [41], Modified Soil Adjusted Vegetation Index (MSAVI) [42], Modified Chlorophyll Absorption Ratio Index (MCARI) [43] and Green Normalised Difference Vegetation Index (GNDVI) [44]. The red, green and blue channels for both cameras were used to compute an extra 4 more indices, namely: Visible Atmospherically Resistant Index (VARI) [45], Visible-band Difference Vegetation Index (VDVI) [46], Normalised Green-Blue Difference Vegetation Index (NGBDI) [47] and Normalised Green-Red Difference Vegetation Index (NGRDI) [48]. The choice for these indices was mostly due to how important the green channel was for measuring reflected vegetation spectra, while also providing more data for FCNNs to start with before modelling complex one-to-one mappings for each pixel.

Calculated indices are concatenated to each image, which results in images for the RedEdge3 camera having 14 bands, whilst the SONY camera has 7. Furthermore, each individual band of data was scaled to a value between 0 and 1.

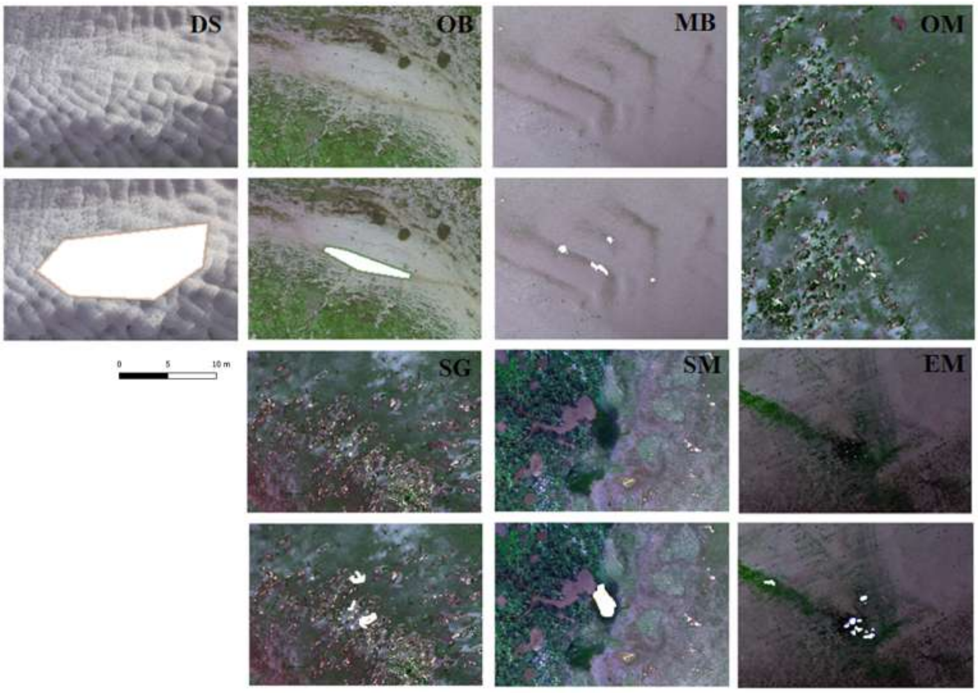

**Figure 3.** Gallery of images and polygons. OM - Other Macroalgae inc. *Fucus*; MB - *Microphytobentos*; EM - *Enteromorpha*; SM - Saltmarsh; SG - Seagrass; DS - Dry Sand; OB - Other Bareground. Images with white polygons are examples of polygons used for modelling.

### *2.5. Fully Convolutional Neural Networks*

Fully Convolutional Neural Networks ([31], [30], [49]) are an extension of traditional CNN architectures ([50], [51]) adapted for semantic segmentation. CNNs are generally comprised of a series of layers that perform feature learning through alternate convolution and pooling operations and a final classification layer. Each convolution and pooling layer transform the input image into higher level abstracted representations. FCNNs can be broken down into two networks: an encoder and a decoder network. The encoder network is identical to a CNN, however the final classification layer is removed. The decoder network applies alternate upsample and convolution operations on feature maps created by the encoder network and a final classification layer with $1 \times 1$ convolution kernels and a softmax function. Network weights and biases are adjusted through gradient descent by computing the crossentropy between network probabilities and encoded values within an segmentation mask.

Figure 4 displays the architecture used for this work. The overall architecture is a U-Net [31] and the encoder network is a ResNet101 [52] pre-trained on ImageNet. Residual learning has proven to surpass very deep neural networks [52] and is a suitable encoder network for the overall U-Net architecture. The decoder network applies a transposed $2 \times 2$ convolution for upsampling while also concatenating feature maps from the encoding stage at appropriate resolutions followed by a final $3 \times 3$ convolution. The final $1 \times 1$ convolution condenses feature maps to have the same number of channels as the total number of classes to be modelled before a softmax transfer function classifies each pixel.

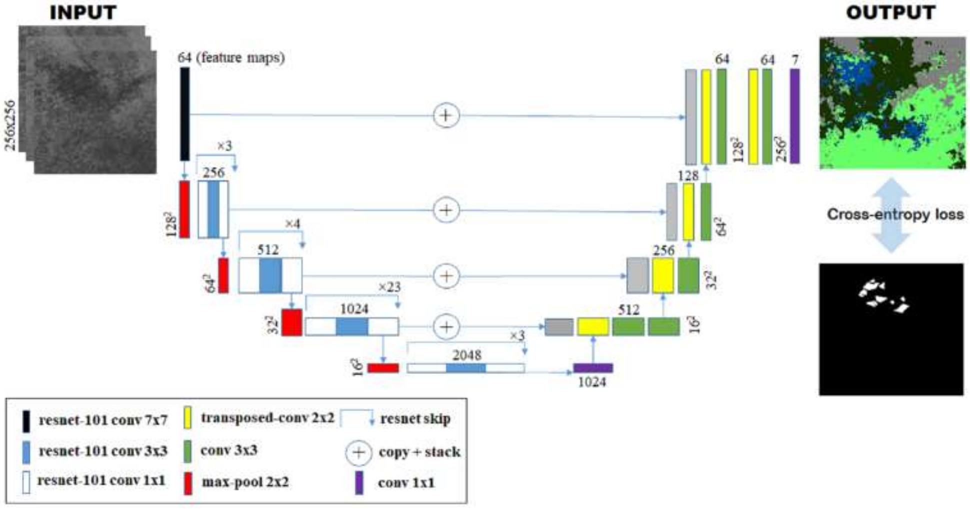

**Figure 4.** U-Net architecture and loss calculation. The input channels are stacked and passed through the network. The encoder network applies repeated convolution and max pooling operations to extract feature maps, while in the decoder network upsamples these and stacks features from the corresponding layer in the encoder path. The output is a segmented map, which is compared with the mask using crossentropy loss. The computed loss is used to train the network, through gradient descent optimisation

For semi-supervised training the Teacher-Student method was used [53]. This approach requires two networks: a teacher and a student model with both models having the same architecture as shown in Figure 4. The student network was updated through gradient descent by adding two loss terms: a supervised loss calculated on labelled pixels of each segmentation map and conversely an unsupervised loss calculated on non-labelled pixels. The teacher network was updated through an exponential moving average of weights within the student network.

#### 2.5.1. Weighted training for FCNNs

Section 2.4 detailed the process to creating segmentation maps from polygons. Both sets of images from each camera had an imbalanced target class distribution. Figure 5 displays the number of labelled pixels per class and also the number of non-labelled pixels for each camera.

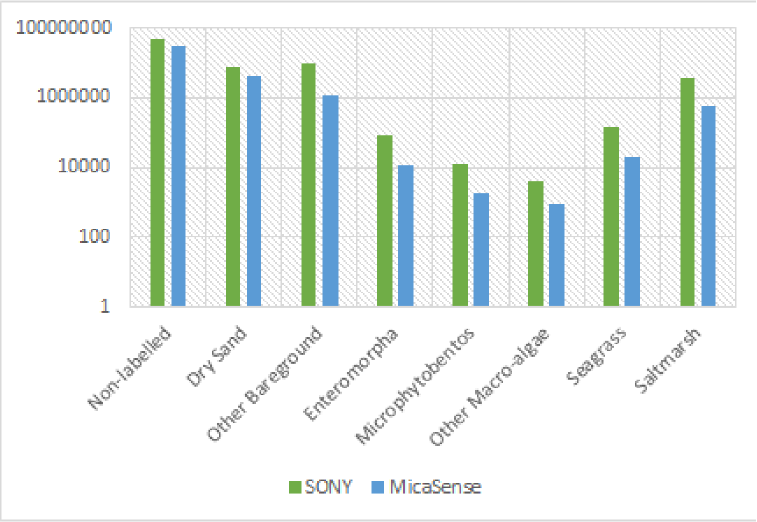

**Figure 5.** Distribution of labelled pixels for each class and non-labelled pixels.

The recorded distribution poses a challenge for classes such as other macro-algae and *Microphytobentos* due to the relative number of labelled pixels in comparison with the remaining target classes. The pixel counts shown in Figure 5 were used to calculate the probability of each class occurring within the training set, and for each class a weight was calculated by taking the inverse for each probability and scaling the weight with respect to the total number of classes.

$$w_i = (p_i.K)^{-1} \tag{1}$$

Where, $w_i$ is $i^{th}$ weight for a given class probability $p_i$ and $K$ is the total number of classes. During FCNN training the supervised loss was scaled with respect to the weights generated in equation 1.

### 2.5.2. Supervised loss

For the supervised loss term, consider $X \in \mathbb{R}^{B\times C\times H\times W}$ and $Y \in \mathbb{Z}^{B\times H\times W}$ to be respectively, a mini-batch of images and corresponding segmentation maps; where $B$, $C$, $H$ and $W$ are respectively, batch size, number of input channels, height and width. Processing a mini-batch with the student network outputs per-pixel scores $\hat{Y} \in \mathbb{R}^{B\times K\times H\times W}$; where $K$ is the number of target classes. The softmax transfer function converts network scores into probabilities by normalising all $K$ scores for each pixel to sum to one.

$$P_k(\mathbf{x}) = \frac{\exp \hat{Y}_k(\mathbf{x})}{\sum_{k'=1}^{K} \exp \hat{Y}_{k'}(\mathbf{x})} \tag{2}$$

Where, $\mathbf{x} \in \Omega$; $\Omega \subseteq \mathbb{Z}^2$ is a pixel location and $P_k(\mathbf{x})$ is the probability for the $k^{th}$ channel at pixel location $\mathbf{x}$, with $\sum_{k'=1}^{K} P_{k'}(\mathbf{x}) = 1$. The negative log-likelihood loss was calculated between segmentation maps and network probabilities.

$$L_s(P, Y) = \begin{cases} 0, \quad \text{if } Y(\mathbf{x}) = -1 \\ -\sum_{k=1}^{K} Y_k(\mathbf{x}) \log(P_k(\mathbf{x})), \\ \qquad\qquad \text{if } Y(\mathbf{x}) \neq -1 \end{cases} \tag{3}$$

For each image, the supervised loss was the sum of all losses for each pixel using eq. 3 and averaged according to the number of labelled pixels within $Y$.

### 2.5.3. Unsupervised loss

Previous work in semi-supervised segmentation details using a Teacher-Student model and advanced data augmentation methods in order to create two images for each network to process ([54], [55]). While this work did not use data augmentation methods, pairs of images were created by using labelled and non-labelled pixels within $Y$.

Similarly to the supervised loss term, a mini-batch of images was processed through both the student and teacher network, respectively producing per-pixel scores $\hat{Y}$ and $\bar{Y}$. Again, pixel scores were converted to probabilities with softmax (eq. 2), $\hat{P}$ and $\bar{P}$, respectively for the student and teacher network. The maximum-likelihood of teacher predictions was used to create pseudo segmentation maps to compute the loss for non-labelled pixels of $Y$. Thus, the unsupervised loss was also calculated similarly to 3 but the negative log-likelihood was computed between predictions from the student model ($\hat{P}$) and a pseudo map ($Y^p$) of pixels that were initially labelled as -1.

$$L_u(\hat{P}, Y^p) = \begin{cases} 0, \quad \text{if } Y(\mathbf{x}) \neq -1 \\ -\sum_{k=1}^{K} Y_k^p(\mathbf{x}) \log(\hat{P}_k(\mathbf{x})), \\ \qquad\qquad \text{if } Y(\mathbf{x}) = -1 \end{cases} \tag{4}$$

For each image, the unsupervised loss was the sum of all losses for each pixel using eq. 4 averaged according to the number of non-labelled pixels within $Y$. The latter loss

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2021 doi:10.20944/preprints202103.0780.v1

was also scaled with respect to the confidence in predictions for the teacher network so that initial optimisation steps focus more on the supervised loss term. Classes with low labelled pixel count would benefit from the unsupervised loss term, as confident teacher predictions can guide the decision boundaries of student models by adding pseudo maps to consider.

#### 2.5.4. Training parameters

Combining both loss terms yields the objective cost used for optimising FCNNs in a semi-supervised setting.

$$L = wL_s + \gamma L_u \tag{5}$$

Where, $L_s$ and $L_u$ are respectively the supervised and unsupervised loss term, with the former being scaled according to the weights computed in 1 and the latter to $\gamma$ which was set to 0.1 for all experiments.

All networks were pre-trained on ImageNet. Networks for each camera were trained for 150 epochs with a batch-size of 16 using Adam optimiser. The learning rate was initially set to 0.001 and reduced by a factor of 10 every 70 epochs of training.

### *2.6. OBIA*

The OBIA method for modelling multiple coastal features was done using eCognition v9.3 [56]. This software has the tools to process the high volume orthomosaics and shape file exports from GISs to create supervised models. Section 2.4 detailed a number of methods used to pre-process the orthomosaics and shape polygons, however the OBIA does not require this.

The first step in OBIA is to process each orthomosaic using a multiscale segmentation to partition the image into segments, also known as image-objects. The segmentation starts with individual pixels and clusters pixels to image-objects using one or more criteria of homogeneity. The subsequent clustering of two adjacent image-objects or image-objects that are a subset of each other were merged together on the following criteria:

$$h = \sum_{c} N(o_c^m - o_c^1) + M(o_c^m - o_c^2) \tag{6}$$

With, $o^1$, $o^2$ and $o^m$ respectively representing the pixel values for object 1, 2 and a candidate virtual merge $m$. $N$ and $M$ are, respectively, the number of total pixels for object 1 and 2. This criteria evaluates the change in homogeneity during fusion of image-objects. If this change exceeds a certain threshold value then the fusion is not performed. In contrast, if the change in image-objects is lower then both candidates are clustered to form a larger segment. The segmentation procedure stops when no further fusions are possible without exceeding the threshold value. In eCognition, this threshold value is a hyper-parameter defined at the start of the process, also known as the scale parameter. The geometry of each shape is defined by two other hyper-parameters: shape and compactness. For this work, the scale parameter is set to 200, the shape to 0.1 and the compactness to 0.5. Figure 6 displays image objects overlaid on top of both orthomosaics.

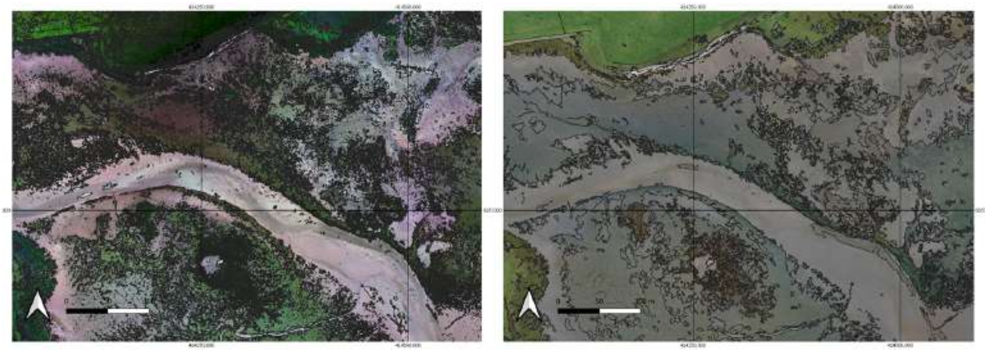

**Figure 6.** Segmented orthomosaics using the multiscale segmentation algorithm. The scale, shape and compactness were respectively 200, 0.1 and 0.5.

Further to this, the polygons (Figure 3) were overlaid on top of image-objects to select the candidate segments for extracting spectral features. Selected image-objects create a database for the in-built Random Forest modeller within eCognition. The spectral features for the RedEdge3 camera were: channel mean and standard deviation, vegetation and soil indices (NDVI, RVI, GNDVI, SAVI), ratios between red/blue, red/green and blue/green image layers and the intensity and saturation components by changing to the HSI colour space. The features for the SONY were the same but the vegetation and soil indices were not added. Once the features and image-objects were selected, the Random Forest modeller produced a number of Decision Trees [57] with each tree being optimised on features using the GINI Index.

## 3. Results

The outputs for both the FCNNs and OBIA were compared with a subset of polygons that were not used for training. Figures 7 and 8 display confusion matrices scoring outputs from each method and camera. The reported accuracy measurements reflect pixel accuracy which is the ratio between pixels that were classified correctly with the total number of labelled pixels within the test set for a given class. A comparison of FCNN models that were trained using only the supervised loss and with both loss terms was also performed.

Overall results for each method and camera can be viewed in Table 1, where precision, recall and F1-score were reported. Precision and recall are metrics that can detail how a classifier performed for each specific class. For instance, low recall can indicate underfit due to a high false negative rate, conversely low precision can indicate overfit due to a high false positive rate. F1-score is the harmonic mean of recall and precision and is a good metric to quantify classifier performance.

$$precision = \frac{TP}{TP + FP} \tag{7}$$

$$recall = \frac{TP}{TP + FN} \tag{8}$$

$$F1 = 2 \times \frac{recall \times precision}{recall + precision} \tag{9}$$

Table 1: Precision, recall and F1 scores for both algorithms on both cameras. DS - Dry Sand; OB - Other bareground; EM - Enteromorpha; MB - Microphytobentos; OM - Other macro-algae; SG - Seagrass; SM - Saltmarsh

| | P | R | F1 | P | R | F1 | P | R | F1 | P | R | F1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DS** | 0.99 | 0.62 | 0.76 | 0.97 | 0.99 | **0.98** | 1.0 | 1.0 | **1.0** | 0.99 | 1.0 | 0.99 |
| **OB** | 0.56 | 0.42 | 0.48 | 0.99 | 0.90 | **0.94** | 0.99 | 0.98 | **0.99** | 0.99 | 0.97 | 0.98 |
| **EM** | 0.73 | 0.95 | 0.83 | 0.91 | 0.91 | **0.91** | 0.25 | 0.97 | **0.40** | 0.18 | 0.57 | 0.27 |
| **MB** | 0.008 | 0.72 | 0.01 | 0.06 | 1.0 | **0.12** | 1.0 | 0.88 | **0.93** | 0.30 | 0.99 | 0.46 |
| **OM** | 0.25 | 0.49 | 0.33 | 0.69 | 0.73 | **0.71** | 0.02 | 0.83 | 0.05 | 0.66 | 0.55 | **0.60** |
| **SG** | 0.67 | 0.95 | **0.78** | 0.64 | 0.73 | 0.68 | 0.64 | 0.93 | **0.76** | 0.27 | 0.93 | 0.42 |
| **SM** | 0.99 | 0.96 | **0.98** | 0.97 | 0.99 | **0.98** | 0.99 | 0.73 | 0.84 | 0.97 | 0.81 | **0.88** |
| | MicaSense: **OBIA** | | | MicaSense: **FCNN** | | | SONY: **OBIA** | | | SONY: **FCNN** | | |

### 3.1. *SONY ILCE-6000 results*

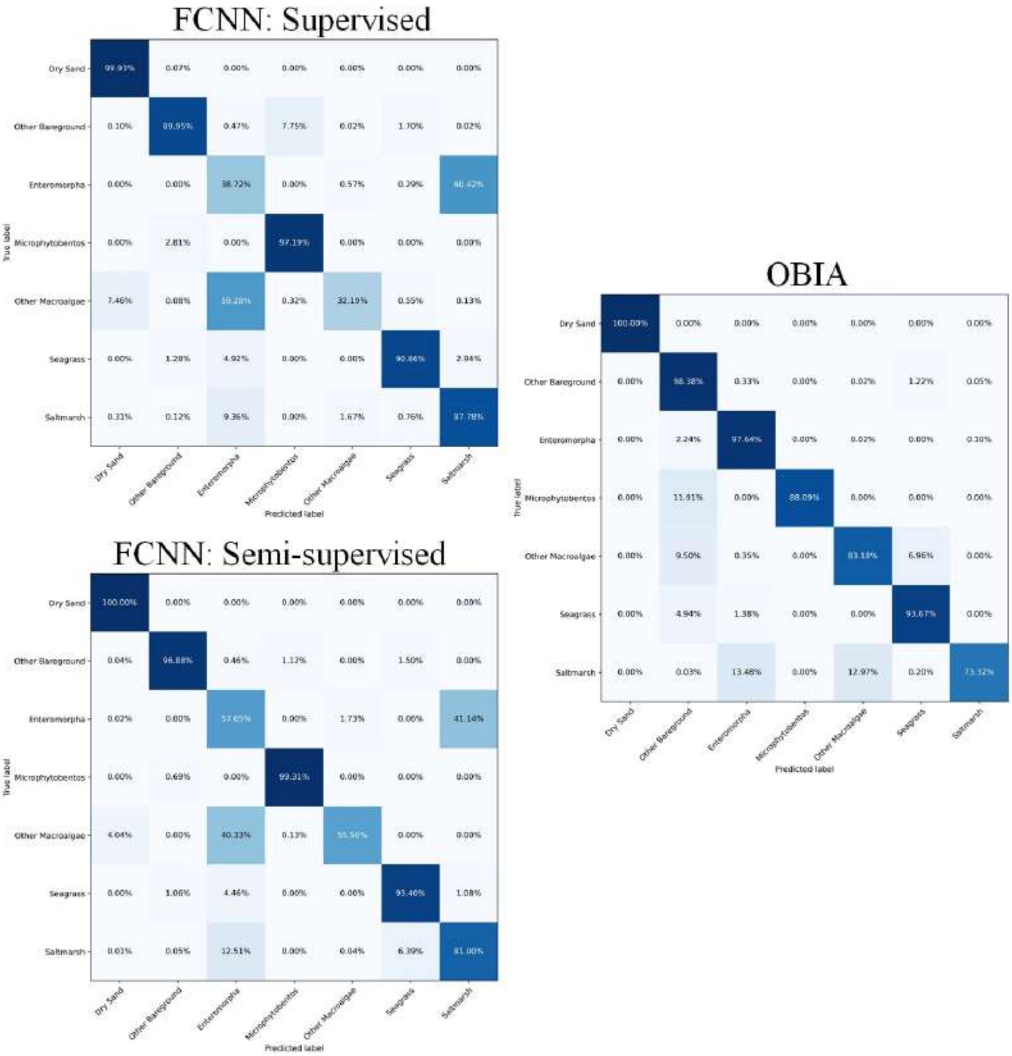

**Figure 7.** Confusion matrices for both methods using the SONY camera.

### *3.2. MicaSense RedEdge3 results*

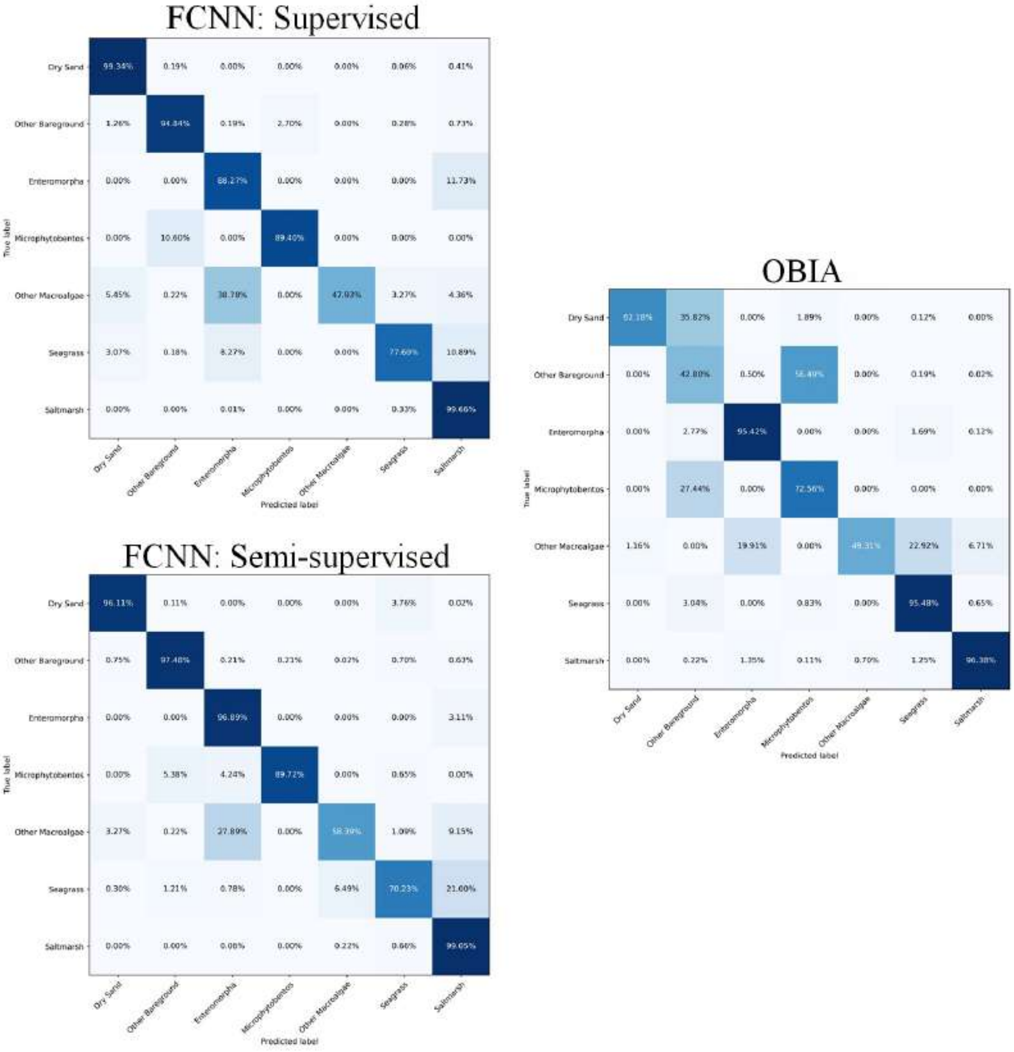

**Figure 8.** Confusion matrices for both methods using the RedEdge3 multispectral camera.

### *3.3. Habitat maps*

Figures 9, 10 and 11 are habitat maps of Budle Bay using each method previously described.

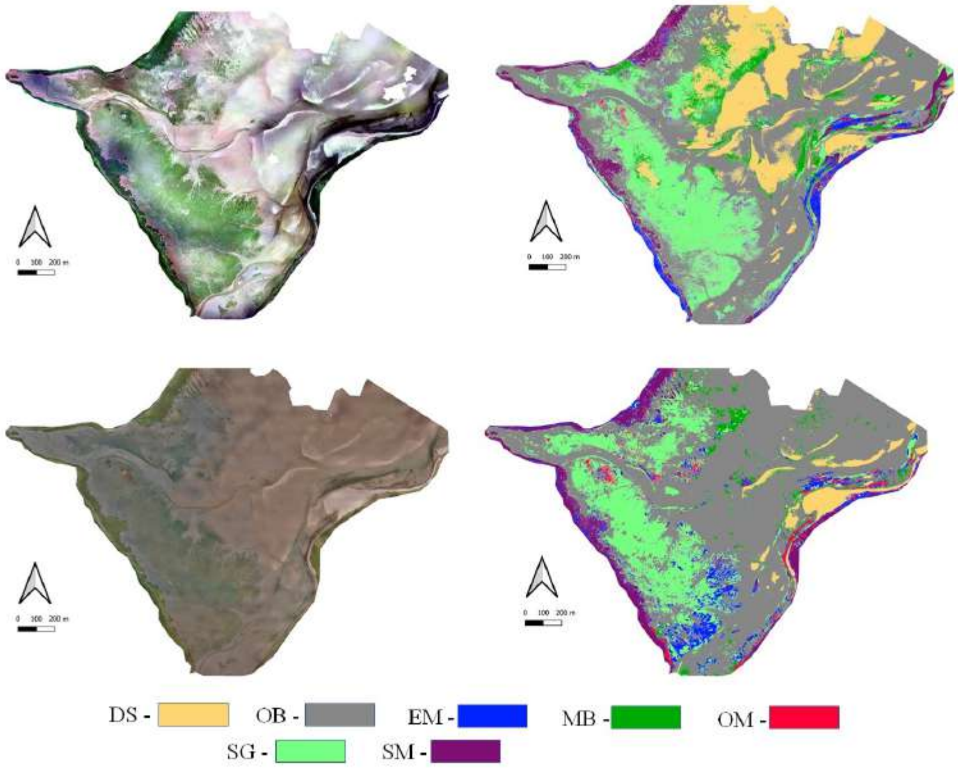

**Figure 9.** Segmented habitat maps for both cameras with OBIA. The top row of images are from the RedEdge3 multispectral camera and the bottom row of images from the SONY camera. Legend: OM - Other Macroalgae inc. *Fucus*; MB - *Microphytobentos*; EM - *Enteromorpha*; SM - Saltmarsh; SG - Seagrass; DS - Dry Sand; OB - Other Bareground.

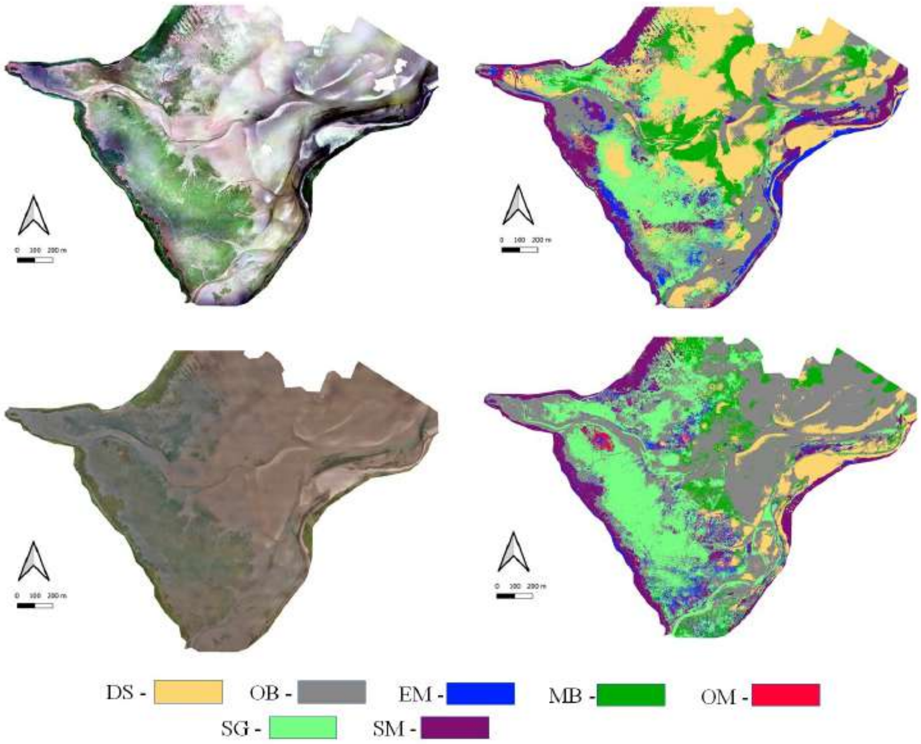

**Figure 10.** Segmented habitat maps for both cameras with FCNNs optimised using only the supervised loss. The top row of images are from the RedEdge3 multispectral camera and the bottom row of images from the SONY camera. Legend: OM - Other Macroalgae inc. *Fucus*; MB - *Microphytobentos*; EM - *Enteromorpha*; SM - Saltmarsh; SG - Seagrass; DS - Dry Sand; OB - Other Bareground.

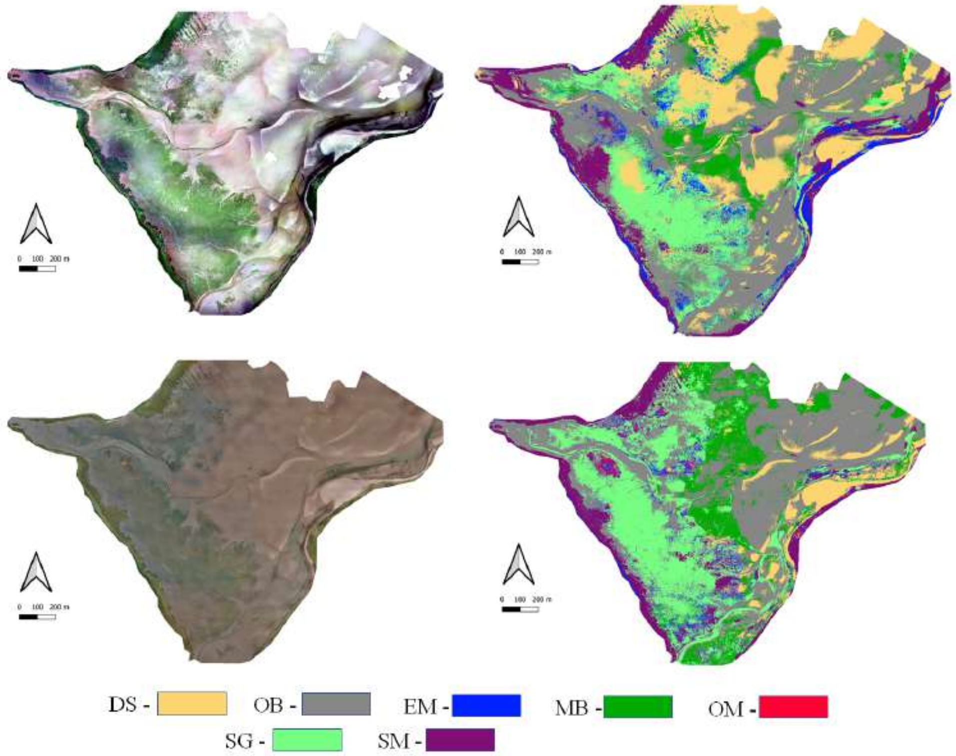

**Figure 11.** Segmented habitat maps for both cameras with FCNNs optimised with both the supervised loss and unsupervised loss. The top row of images are from the RedEdge3 multispectral camera and the bottom row of images from the SONY camera. Legend: OM - Other Macroalgae inc. *Fucus*; MB - *Microphytobentos*; EM - *Enteromorpha*; SM - Saltmarsh; SG - Seagrass; DS - Dry Sand; OB - Other Bareground.

## 4. Discussion

The initial analysis of results was done for each individual camera based off pixel accuracy and an overall discussion for both cameras and methods was described using precision, recall and F1-score.

### *4.1. SONY ILCE-6000 analysis*

The results for the SONY camera in terms of average pixel accuracy across all classes were better with OBIA than FCNNs.

Predictions with the OBIA method had an average pixel accuracy of 90.6%. Classes related to sediment had scores of 100% and 98.38%, respectively for dry sand and other bareground. This method also performed well for all algal classes listed in 2.3 and in particular predictions for other macro-algae scored considerably higher with OBIA than with FCNNs. As mentioned previously, this class in particular had the least amount of labelled pixels (Figure 5) which posed a challenge for FCNNs models. Algal classes scored 97.6%, 88.09% and 83.18%, respectively for *Enteromorpha*, *Microphytobentos* and other macro-algae (inc. *Fucus*). Seagrass predictions were found to score 93.67% and saltmarsh was the worst performing class for the OBIA with 73.32%.

FCNNs in either supervised and semi-supervised training yielded an average class accuracy of 76.79% and 83.3%, respectively. Both approaches to training FCNNs had comparable scores to OBIA with the exception to *Enteromorpha* and other macro-algae, which respectively scored 38.72% and 32.29% for supervised training and 57.05% and 55.90% for semi-supervised training. Other macro-algae was often miss classified as *Enteromorpha* another algal class within the dataset, while *Enteromorpha* was often predicted as saltmarsh. The addition of the unsupervised loss term for semi-supervised training helped increase the pixel accuracy for other macro-algae, supporting the initial

doi:10.20944/preprints202103.0780.v1

premise that target classes with relatively low pixel label counts can benefit from the semi-supervised approach described in section 2.5.

### 4.2. MicaSense RedEdge3 analysis

The results for the RedEdge3 camera in terms of average pixel accuracy across all classes were mostly better with FCNNs in both training modes than OBIA.

The OBIA method had an average pixel accuracy of 73.44%. Sediment classes such as dry sand and other bareground scored 63.18% and 42.80%, respectively, with *Microphytobentos* being classified 56.49% as other bareground. *Microphytobentos* consists of a unicellular eukaryotic algae and cyanobacteria that grow within the upper millimeters of illuminated sediments, typically appearing only as a subtle greenish shading [58], and since other bareground includes wet sand, both classes were interchangeably miss classified. Algal classes scored 93.42%, 72.54% and and 49.31%, respectively for *Enteromorpha*, *Microphytobentos* and other macro-algae (inc. *Fucus*). The remaining vegetation classes of seagrass and saltmarsh both presented high scores of 95.48% and 96.38, respectively, and in particular seagrass predictions were more accurate with OBIA than FCNNs. Figure 6 displays image-objects generated as a result of MSS with the number of image-objects present for the RedEdge3 camera being greater than objects generated using the SONY camera. This indicates that the scale parameter should be reduced for the multispectral camera since Figure 7 suggests that OBIA with the SONY camera was a suitable method for this work.

FCNNs in either supervised and semi-supervised training yielded an average class accuracy of 85.27% and 88.44%, respectively. Both models had robust scores for sediment scoring above 95% in pixel accuracy. Again, results between supervised and semi-supervised models for algal classes note an increase in performance when the unsupervised loss term was added to the training algorithm. This also supports the premise that an unsupervised loss term aids FCNNs with target classes that had a low number of labelled pixels relative to the remaining classes. This particularly holds true for classes of *Microphytobenthos* and other macro-algae which benefited the most from a semi-supervised training mode. Predictions for saltmarsh were comparably equal to OBIA, but seagrass was often miss classified as saltmarsh and/or *Enteromorpha*.

### 4.3. Overall analysis

Results for both methods were collated for each camera in Table 1. Note that scores for FCNNs depict models trained in semi-supervision.

Table 1 shows that F1-scores for the SONY camera favoured the OBIA approach. While sediment classes had equal performance to FCNNs, predictions for algal classes of *Enteromorpha* and *Microphytobentos* scored better with OBIA. In particular, the latter class was far better due to FCNNs recording low precision which could suggest some degree of overfit. This conclusion can be supported by large areas that were predicted as *Microphytobentos* in Figures 10 and 11 for the SONY camera due to a high false-positive rate. Other macro-algae was found to have a better F1-score with FCNNs than OBIA which contradicts the initial analysis in Figure 7. While seagrass scores for pixel accuracy were equal for both methods (Figure 7), the precision for FCNNs was lower than OBIA, again suggesting a high false-positive rate for FCNNs.

F1-scores for the RedEdge3 camera show predictions with FCNNs were robust for classes related to sediment, *Enteromorpha* and saltmarsh. The latter classes all scored above 0.90 for F1-score, suggesting that FCNNs avoid high false-positive and false-negative rates for the latter classes. This was not the case for *Microphtobentos* predictions, where Figure 8 suggests high accuracy but Table 1 noted low precision and high recall, which suggests some degree of overfit. Again, this conclusion can be supported by large areas that were predicted as *Microphytobentos* in Figures 10 and 11 for the RedEdge3. Other macro-algae (inc. *Fucus*) was found to be a problematic class to model for OBIA due to the small number of labelled pixels relative to the rest of the dataset. But this was

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2021 doi:10.20944/preprints202103.0780.v1

not the case for FCNNs where a combination of weighted trained and semi-supervision through an unsupervised loss helped to guide decision boundaries towards accurate predictions. Lastly, seagrass predictions were found to be more accurate with OBIA due to higher scores in recall, suggesting that FCNNs were more likely to miss classify seagrass polygons. This can be loosely noted by comparing seagrass predictions in Figure 9 with corresponding areas that were not predicted as seagrass in Figures 10 and 11.

Generally for seagrass mapping, both methods for both cameras exhibit confident predictions for seagrass, with the worse combination of camera and approach being FCNNs with the SONY camera. In fact, both cameras with FCNNs performed worse with regards to seagrass, but the SONY camera was due to a high false-positive rate and the RedEdge3 was due to a high false negative rate. Another final consideration was the pre-process necessary to create segmentation maps for FCNN training, a process that is not required in Geo-OBIA software such as eCognition.

Comparing the results of seagrass mapping through FCNNs with previous attempts found use of the same metrics stated in section 3. However, most of these studies ([59], [60], [61]) were mainly concerned with seagrass meadows instead of intertidal seagrass. FCNNs have been used for mapping intertidal macro-algae [62] with reported average accuracies for a 5 class problem to be 91.19%. However, this work considers mapping intertidal macro-algae, seagrass and sediment features at a coarser resolution. In fact, it is to the authors knowledge that this is the first use of FCNNs for intertidal seagrass mapping. Previous attempts with OBIA were found to be successful ([16], [19]) with the former using the root mean-square error to quantify results for a 5 class problem and the latter using accuracy in classified image-objects instead of per-pixel.

#### 4.3.1. Habitat map discussion

The confusion matrices in Figures 7 and 8 and F1-scores in Table 1 detail objective scores based off data from the in-situ survey, and while these polygons provide a form of preliminary analysis for each method, correlations between confusion matrices and predicted maps can be found.

For instance, the OBIA method using the multispectral camera for seagrass classifications was found to be more accurate than FCNNs; interestingly all FCNNs trained and tested on the latter camera did not predict seagrass along the North Western part of the site (area covered in Figure 6). While it is not possible to quantify which method was correct without surveying the site again, the confidence in seagrass predictions for OBIA along with FCNNs predicting bareground sediment instead of vegetation can lead to users being more confident with OBIA for seagrass mapping.

Another key difference between each method were the areas to which *Microphytobentos* was predicted while also maintaining confident scores with respect to pixel accuracy. As mentioned, *Microphytobentos* consists of a unicellular algae specie that grow within the upper millimeters of illuminated sediments, which could suggest that the total area covered by *Microphytobentos* should not be as large as depicted in maps processed by FCNNs. Scores for precision shown in Table 1 support this claim due to low precision indicating some degree of overfit which in turn explain large areas predicted as *Microphythobentos*. The only exception to this being the OBIA with the SONY camera.

The remaining classes were similar in areas to which they covered within habitat maps.

## 5. Conclusion

In this work we show that FCNNs can model high resolution aerial imagery from a small set of polygons that were converted to segmentation maps for training. Each FCNN was evaluated in two training modes, supervised and semi-supervised, with results indicating that semi-supervision help models with target classes that have a small number of labelled pixels. A prospect that may be of benefit for studies where in-situ surveying is an expensive effort to conduct.

This work also showed that OBIA continues to be a robust approach for monitoring multiple coastal features in high resolution imagery. In particular, both cameras with the OBIA were found to be more accurate over FCNNs in predicting seagrass. However, as noted in section 3 results with OBIA were highly dependant on the initial parameters used for MSS, with the scale parameter being critical for image-object creation.

The study site and problem formation described in section 2.3 combined for a complex problem. This in turn can make confidence in seagrass predictions decrease as ambiguity over multiple vegetation classes increases. OBIA was found to overcome this with both cameras accurately predicting seagrass polygons while maintaining relatively high precision when compared to FCNNs. On the other hand, FCNNs were found to be more accurate to classify algal classes, and in particular for other macro-algae which had the least number of labelled pixels. Therefore, while this work shows that OBIA is a suitable method for intertidal seagrass mapping, other applications within remote sensing for coastal monitoring with restricted access to in-situ data can leverage semi-supervision.

**Author Contributions:** Conceptualization, B.H., M.M., J.B., T.D.; methodology, B.H., R.A., G.F.; software, B.H., R.A. G.F.; data curation, B.H., J.B. T.D.; writing—original draft preparation, B.H.; writing—review and editing, B.H., M.M., J.B., T.D., R.A.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Cefas, Cefas Technology and the Natural Environmental Research Council through the NEXUSS CDT, grant number NE/RO12156/1, titled "Blue eyes: New tools for monitoring coastal environments using remotely piloted aircraft and machine learning"

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

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| RS | Remote Sensing |
| FCNN | Fully Convolutional Neural Network |
| MSS | Multi-Scale Segmentation |
| OBIA | Object-Based Image Analysis |

## References

1. Fonseca, M.S.; Zieman, J.C.; Thayer, G.W.; Fisher, J.S. The role of current velocity in structuring eelgrass (Zostera marina L.) meadows. *Estuarine, Coastal and Shelf Science* **1983**, *17*, 367–380.
2. Fonseca, M.S.; Bell, S.S. Influence of physical setting on seagrass landscapes near Beaufort, North Carolina, USA. *Marine Ecology Progress Series* **1998**, *171*, 109–121.
3. Gera, A.; Pagès, J.F.; Romero, J.; Alcoverro, T. Combined effects of fragmentation and herbivory on Posidonia oceanica seagrass ecosystems. *Journal of ecology* **2013**, *101*, 1053–1061.
4. Pu, R.; Bell, S.; Meyer, C. Mapping and assessing seagrass bed changes in Central Florida's west coast using multitemporal Landsat TM imagery. *Estuarine, Coastal and Shelf Science* **2014**, *149*, 68–79.
5. Short, F.T.; Wyllie-Echeverria, S. Natural and human-induced disturbance of seagrasses. *Environmental conservation* **1996**, pp. 17–27.
6. Marbà, N.; Duarte, C.M. Mediterranean warming triggers seagrass (Posidonia oceanica) shoot mortality. *Global Change Biology* **2010**, *16*, 2366–2375.
7. Duarte, C.M. The future of seagrass meadows. *Environmental conservation* **2002**, pp. 192–206.
8. McGlathery, K.J.; Reynolds, L.K.; Cole, L.W.; Orth, R.J.; Marion, S.R.; Schwarzschild, A. Recovery trajectories during state change from bare sediment to eelgrass dominance. *Marine Ecology Progress Series* **2012**, *448*, 209–221.

---

9. Lamb, J.B.; Van De Water, J.A.; Bourne, D.G.; Altier, C.; Hein, M.Y.; Fiorenza, E.A.; Abu, N.; Jompa, J.; Harvell, C.D. Seagrass ecosystems reduce exposure to bacterial pathogens of humans, fishes, and invertebrates. *Science* **2017**, *355*, 731–733.
10. Fourqurean, J.W.; Duarte, C.M.; Kennedy, H.; Marbà, N.; Holmer, M.; Mateo, M.A.; Apostolaki, E.T.; Kendrick, G.A.; Krause-Jensen, D.; McGlathery, K.J.; others. Seagrass ecosystems as a globally significant carbon stock. *Nature geoscience* **2012**, *5*, 505–509.
11. Macreadie, P.; Baird, M.; Trevathan-Tackett, S.; Larkum, A.; Ralph, P. Quantifying and modelling the carbon sequestration capacity of seagrass meadows–a critical assessment. *Marine pollution bulletin* **2014**, *83*, 430–439.
12. Dennison, W.C.; Orth, R.J.; Moore, K.A.; Stevenson, J.C.; Carter, V.; Kollar, S.; Bergstrom, P.W.; Batiuk, R.A. Assessing water quality with submersed aquatic vegetation: habitat requirements as barometers of Chesapeake Bay health. *BioScience* **1993**, *43*, 86–94.
13. Waycott, M.; Duarte, C.M.; Carruthers, T.J.; Orth, R.J.; Dennison, W.C.; Olyarnik, S.; Calladine, A.; Fourqurean, J.W.; Heck, K.L.; Hughes, A.R.; others. Accelerating loss of seagrasses across the globe threatens coastal ecosystems. *Proceedings of the national academy of sciences* **2009**, *106*, 12377–12381.
14. Richards, J.A.; Richards, J. *Remote sensing digital image analysis*; Vol. 3, Springer, 1999.
15. Anderson, K.; Gaston, K.J. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment* **2013**, *11*, 138–146.
16. Duffy, J.P.; Pratt, L.; Anderson, K.; Land, P.E.; Shutler, J.D. Spatial assessment of intertidal seagrass meadows using optical imaging systems and a lightweight drone. *Estuarine, Coastal and Shelf Science* **2018**, *200*, 169–180.
17. Turner, D.; Lucieer, A.; Watson, C. An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (UAV) imagery, based on structure from motion (SfM) point clouds. *Remote sensing* **2012**, *4*, 1392–1410.
18. Collin, A.; Dubois, S.; Ramambason, C.; Etienne, S. Very high-resolution mapping of emerging biogenic reefs using airborne optical imagery and neural network: the honeycomb worm (Sabellaria alveolata) case study. *International Journal of Remote Sensing* **2018**, *39*, 5660–5675.
19. Ventura, D.; Bonifazi, A.; Gravina, M.F.; Belluscio, A.; Ardizzone, G. Mapping and classification of ecologically sensitive marine habitats using unmanned aerial vehicle (UAV) imagery and object-based image analysis (OBIA). *Remote Sensing* **2018**, *10*, 1331.
20. Rossiter, T.; Furey, T.; McCarthy, T.; Stengel, D.B. UAV-mounted hyperspectral mapping of intertidal macroalgae. *Estuarine, Coastal and Shelf Science* **2020**, *242*, 106789.
21. Foody, G.M. Status of land cover classification accuracy assessment. *Remote sensing of environment* **2002**, *80*, 185–201.
22. Blaschke, T. Object based image analysis for remote sensing. *ISPRS journal of photogrammetry and remote sensing* **2010**, *65*, 2–16.
23. Su, W.; Li, J.; Chen, Y.; Liu, Z.; Zhang, J.; Low, T.M.; Suppiah, I.; Hashim, S.A.M. Textural and local spatial statistics for the object-oriented classification of urban areas using high resolution imagery. *International journal of remote sensing* **2008**, *29*, 3105–3117.
24. Flanders, D.; Hall-Beyer, M.; Pereverzoff, J. Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction. *Canadian Journal of Remote Sensing* **2003**, *29*, 441–452.
25. Butler, J.D.; Purkis, S.J.; Yousif, R.; Al-Shaikh, I.; Warren, C. A high-resolution remotely sensed benthic habitat map of the Qatari coastal zone. *Marine Pollution Bulletin* **2020**, *160*, 111634.
26. Husson, E.; Ecke, F.; Reese, H. Comparison of manual mapping and automated object-based image analysis of non-submerged aquatic vegetation from very-high-resolution UAS images. *Remote Sensing* **2016**, *8*, 724.
27. Purkis, S.J.; Gleason, A.C.; Purkis, C.R.; Dempsey, A.C.; Renaud, P.G.; Faisal, M.; Saul, S.; Kerr, J.M. High-resolution habitat and bathymetry maps for 65,000 sq. km of Earth's remotest coral reefs. *Coral Reefs* **2019**, *38*, 467–488.
28. Rasuly, A.; Naghdifar, R.; Rasoli, M. Monitoring of Caspian Sea coastline changes using object-oriented techniques. *Procedia Environmental Sciences* **2010**, *2*, 416–426.
29. Schmidt, K.; Skidmore, A.; Kloosterman, E.; Van Oosten, H.; Kumar, L.; Janssen, J. Mapping coastal vegetation using an expert system and hyperspectral imagery. *Photogrammetric Engineering & Remote Sensing* **2004**, *70*, 703–715.

30. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.
31. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.
32. Cordts, M.; Omran, M.; Ramos, S.; Scharwächter, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The cityscapes dataset. CVPR Workshop on the Future of Datasets in Vision, 2015, Vol. 2.
33. Everingham, M.; Eslami, S.A.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes challenge: A retrospective. *International journal of computer vision* **2015**, *111*, 98–136.
34. Ladle, M. The Haustoriidae (Amphipoda) of Budle Bay, Northumberland. *Crustaceana* **1975**, *28*, 37–47.
35. Meyer, A. An investigation into certain aspects of the ecology of Fenham flats and budle bay, Northumberland. PhD thesis, Durham University, 1973.
36. Olive, P. Management of the exploitation of the lugworm Arenicola marina and the ragworm Nereis virens (Polychaeta) in conservation areas. *Aquatic Conservation: Marine and Freshwater Ecosystems* **1993**, *3*, 1–24.
37. Agisoft, L. Agisoft metashape user manual, Professional edition, Version 1.5. *Agisoft LLC, St. Petersburg, Russia, from https://www. agisoft. com/pdf/metashape-pro_1_5_en. pdf, accessed June* **2018**, *2*, 2019.
38. Cunliffe, A.M.; Brazier, R.E.; Anderson, K. Ultra-fine grain landscape-scale quantification of dryland vegetation structure with drone-acquired structure-from-motion photogrammetry. *Remote Sensing of Environment* **2016**, *183*, 129–143.
39. Xue, J.; Su, B. Significant remote sensing vegetation indices: A review of developments and applications. *Journal of sensors* **2017**, *2017*.
40. Rouse, J.; Haas, R.H.; Schell, J.A.; Deering, D.W.; others. Monitoring vegetation systems in the Great Plains with ERTS. *NASA special publication* **1974**, *351*, 309.
41. Ren-hua, Z.; Rao, N.; Liao, K. Approach for a vegetation index resistant to atmospheric effect. *Journal of Integrative Plant Biology* **1996**, *38*.
42. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote sensing of environment* **1994**, *48*, 119–126.
43. Daughtry, C.S.; Walthall, C.; Kim, M.; De Colstoun, E.B.; McMurtrey Iii, J. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote sensing of Environment* **2000**, *74*, 229–239.
44. Louhaichi, M.; Borman, M.M.; Johnson, D.E. Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto International* **2001**, *16*, 65–70.
45. Gitelson, A.A.; Kaufman, Y.J.; Stark, R.; Rundquist, D. Novel algorithms for remote estimation of vegetation fraction. *Remote sensing of Environment* **2002**, *80*, 76–87.
46. Xiaoqin, W.; Miaomiao, W.; Shaoqiang, W.; Yundong, W. Extraction of vegetation information from visible unmanned aerial vehicle images. *Transactions of the Chinese Society of Agricultural Engineering* **2015**, *31*.
47. Verrelst, J.; Schaepman, M.E.; Koetz, B.; Kneubühler, M. Angular sensitivity analysis of vegetation indices derived from CHRIS/PROBA data. *Remote Sensing of Environment* **2008**, *112*, 2341–2353.
48. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment* **1979**, *8*, 127–150.
49. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the European conference on computer vision (ECCV), 2018, pp. 801–818.
50. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **1998**, *86*, 2278–2324.
51. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Communications of the ACM* **2017**, *60*, 84–90.
52. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
53. Tarvainen, A.; Valpola, H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *arXiv preprint arXiv:1703.01780* **2017**.

54. French, G.; Laine, S.; Aila, T.; Mackiewicz, M.; Finlayson, G. Semi-supervised semantic segmentation needs strong, varied perturbations. British Machine Vision Conference, 2020, number 31.
55. Olsson, V.; Tranheden, W.; Pinto, J.; Svensson, L. Classmix: Segmentation-based data augmentation for semi-supervised learning. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 1369–1378.
56. Nussbaum, S.; Menz, G. eCognition image analysis software. In *Object-based image analysis and treaty verification*; Springer, 2008; pp. 29–39.
57. Quinlan, J.R. Induction of decision trees. *Machine learning* **1986**, *1*, 81–106.
58. MacIntyre, H.L.; Geider, R.J.; Miller, D.C. Microphytobenthos: the ecological role of the "secret garden" of unvegetated, shallow-water marine habitats. I. Distribution, abundance and primary production. *Estuaries* **1996**, *19*, 186–201.
59. Reus, G.; Möller, T.; Jäger, J.; Schultz, S.T.; Kruschel, C.; Hasenauer, J.; Wolff, V.; Fricke-Neuderth, K. Looking for seagrass: Deep learning for visual coverage estimation. 2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO). IEEE, 2018, pp. 1–6.
60. Weidmann, F.; Jäger, J.; Reus, G.; Schultz, S.T.; Kruschel, C.; Wolff, V.; Fricke-Neuderth, K. A closer look at seagrass meadows: Semantic segmentation for visual coverage estimation. OCEANS 2019-Marseille. IEEE, 2019, pp. 1–6.
61. Yamakita, T.; Sodeyama, F.; Whanpetch, N.; Watanabe, K.; Nakaoka, M. Application of deep learning techniques for determining the spatial extent and classification of seagrass beds, Trang, Thailand. *Botanica marina* **2019**, *62*, 291–307.
62. Balado, J.; Olabarria, C.; Martínez-Sánchez, J.; Rodríguez-Pérez, J.R.; Pedro, A. Semantic segmentation of major macroalgae in coastal environments using high-resolution ground imagery and deep learning. *International Journal of Remote Sensing* **2021**, *42*, 1785–1800.