# Coastline Detection in Satellite Imagery: A Deep Learning Approach on New Benchmark Data

Catherine Seale[a,b,*], Thomas Redfern[a], Paul Chatfield[a], Chunbo Luo[b], Kari Dempsey[a]

[a] *UK Hydrographic Office, Taunton, TA1 2DN, Somerset, United Kingdom*
[b] *Institute for Data Science and Artificial Intelligence, University of Exeter, Exeter, EX4 4PY, Devon, United Kingdom*

## Abstract

Detailed and up-to-date coastline morphology data underpins our understanding of coastline change over time. The development of an automated and scalable coastline extraction methodology from satellite imagery is currently limited by the low availability of open, globally distributed and diverse labelled data with which to develop and benchmark techniques. Therefore, in this study we present the Sentinel-2 Water Edges Dataset (SWED), a new and bespoke labelled image dataset for the development and bench-marking of techniques for the automated extraction of coastline morphology data from Sentinel-2 images. Composed of 16 labelled training Sentinel-2 scenes, and 98 test label-image pairs, SWED is globally distributed and contains examples of many different coastline types and natural and anthropogenic coastline features.

To provide a baseline of model performance against SWED we train and test four convolutional neural network models, based on the U-Net model architecture. Models are optimised using Categorical Cross-entropy Loss, Sørensen–Dice Loss and two novel loss functions we present for the focusing of

*Corresponding author.
*Email address:* catherineseale@gmail.com (Catherine Seale)

model training attention to the boundary between land and water. Through a hybrid quantitative and qualitative model assessment process we demonstrate that the model trained using our novel Sobel-edge loss function has greater sensitivity to fine-scale, narrow coastline features whilst possessing near top quantitative performance demonstrated by Categorical Cross-entropy.

The SWED dataset is published openly for use by the remote sensing and machine learning communities, whilst the Sobel-edge loss is available for use in machine learning applications where sensitivity to boundary features is important.

*Keywords:* Automated coastline extraction, Sentinel-2 satellite imagery, Deep Learning, Machine Learning, Labelled data, Loss Function

*PACS:* 0000, 1111

*2000 MSC:* 0000, 1111

---

## 1. Introduction

Coastal regions hold significant environmental, societal and economic value (Wyles et al., 2019). Throughout history human populations have been attracted to settle in coastal areas, due to the fertile soils, abundant food and opportunities for transport and trade in these regions (Edmonds et al., 2020). This is still the case today, with 41% of the global population living within 100 km of the coastline (Martinez et al., 2007). According to IPCC reports, coastal regions are particularly sensitive to the impacts of climate change, and risks to these areas, from both natural and anthropogenic drivers, threaten both human populations and the ecosystems they rely upon (Wong et al., 2014). Detailed and up-to-date coastline morphology data, defined here as the form or shape of the boundary between land and water, underpins our understanding of coastline change and our ability to manage its impacts over time (Burningham and French, 2017).

There are two main methodological sources of coastline morphology data, in-situ measurements (Kuschnerus et al., 2021) and observations from remote sensing technologies (Zhu et al., 2021). In-situ profiling or surveying provides the most precise results but is only practicable for small regions. The labour and costs involved render in-situ methods unfeasible for mapping extensive areas or for repeated analyses. Indeed, in-situ measurement may be impossible if a study area is remote, treacherous or inaccessible. Likewise, remotely sensed data or imagery captured by piloted flight or drone are expensive to collect and impractical to apply at scale. In consequence, the dominant approaches to mapping coastlines use satellite remote sensing imagery and image processing techniques (Toure et al., 2019). The benefits are numerous; satellites provide a rich time series of images as they revisit an area every few days, they allow measurement without having to travel to an area, and different instrumentation is available for a variety of use cases. As such, satellite remote sensing is the only source of data by which we might realistically continuously observe global coastline morphology. Notable examples of free and open-access global satellite data are the European Space Agency's Copernicus Programme (European Commission, 2013), and the Landsat archive from NASA and the U.S. Geological Survey (Loveland and Dwyer, 2012). These multi-petabyte datasets enable monitoring of the Earth's surface over time (Li and Gong, 2016), presenting opportunities to study the world's natural and anthropogenic environments at scale. Therefore global, repeatable analysis of coastline morphology will be possible, if a sufficiently automated, scalable and accurate method is devised for processing coastal satellite images into coastline morphology data (Parente et al., 2019).

This opportunity does however come with intrinsic limitations. Firstly, the temporal record available for a satellite is fixed to the lifetime of its operation, limiting potential for longer-term analysis. For example, data collection by the European Space Agency's Sentinel-2 mission commenced in 2015, therefore analysis prior to this date is not possible using this sensor alone. A longer archive is offered by the Landsat mission, with 30m resolution imagery available from 1999, however, this is of coarser spatial resolution than Sentinel-2. Admitting the limitations of spatial resolution and temporal record with respect to global satellite data for the time being, the data collected by these missions are proving suitable for study of coastlines. At the time of writing, a small number of related analyses have been performed on global satellite data. These examples include monitoring the extent of sandy beaches (Luijendijk et al., 2018), tidal flats (Murray et al., 2018) and mangrove (Bunting et al., 2018). But as yet, the lack of a viable, scalable global method to observe and monitor coastline morphology still limits our understanding of coastal zones. At present, the majority of studies performing coastline morphology detection restrict their focus to small local regions (Uddin et al., 2020), with one notable exception from Bishop-Taylor et al. (2019) that successfully extends detection to the Australian continent.

Local-scale, automated approaches to coastline morphology detection within satellite imagery fall into two main categories. The first category consists of edge detection methods, which aim to detect the coastline as a linear feature within an image (Karantzalos et al., 2002; Liu and Jezek, 2004; Heene and Gautama, 2000; Paravolidakis et al., 2018; Klinger et al., 2012). Edge detection methods are straightforward to compute and do not need any specialist knowledge of a coastline's specific characteristics, but are often sensitive to noise, require manual intervention and are not generally recommended for use over large geographic areas (Toure et al., 2019). The second category are segmentation methods, which aim to classify image pixels into regions, or 'segments', with the coastline defined as the boundary between water vs. non-water segments (Cao et al., 2020).

Owing to the widespread availability and reduced entry cost to powerful computing, machine learning (ML) and deep learning (DL) approaches to image segmentation tasks are now commonplace in the analysis of satellite imagery (Kattenborn et al., 2021). Convolutional neural networks (CNNs) are the current state-of-the-art in the field of image segmentation (Sultana et al., 2020), and while examples of CNN in the marine domain are few in comparison to terrestrial applications (Yuan et al., 2020), there are a small number of studies that apply CNNs to the task of coastline detection. Cheng et al. (2016) and Li et al. (2018) use CNNs to delineate sea-land boundaries while Vos et al. (2019) detect sandy shorelines at five beach locations using a multi-layer perceptron, applying their analysis at scale by leveraging archives of Landsat and Sentinel-2 imagery made available via the Google Earth Engine (GEE) platform.

Development of a single ML or DL model with the capability of accurately detecting coastline morphology from satellite imagery anywhere in the world is a complex and non-trivial task, due to the diversity of coastline features world-wide. Phenomena such as waves, variable geology and water turbidity all contribute to variety in the appearance of the coastline within satellite imagery (Toure et al., 2019). Poor spatio-temporal generalisation and scalability is a common issue for all DL approaches to satellite and aerial imagery analysis (Wang et al., 2017) and coastlines in particular are tremendously diverse, having both natural boundaries, such as beaches, mangroves or cliffs, and anthropogenic boundaries, such as piers, pipelines and harbours. Of particular difficulty for CNN models are narrow, linear features of a size close to the native image resolution, for example, a 10-metre-wide pier protruding into the sea is represented in 10 metre resolution imagery with only one pixels width (Cheng et al., 2016). Standard loss functions used for training CNN models, including cross-entropy loss and Sørensen-Dice loss, influence model training using an aggregate of per-pixel error, computed discretely at each pixel location. In this way, each pixel is treated with equal weighting, regardless of whether it is part of a class boundary. Ignoring the relationship between neighbouring pixels during optimisation limits a model's sensitivity to those features that are of small size, linear, narrow or at the boundary of segmentation targets, making important coastal features, such as piers, difficult to detect (Cheng et al., 2016). Insensitivity of detectors to fine-grained narrow features is not an issue limited to coastlines, rather any narrow linear feature, and specialised methods for detecting other linear features such as rivers (Yang et al., 2015) and roads (Oehmcke et al., 2019) have been developed.

From this relatively small body of research, it is not possible to identify a state-of-the-art method for coastline morphology detection with any certainty. There is no algorithm that can be used regardless of geography or coastline type, and with sensitivity to coastal features that are narrow or small at image resolution. Current methods have focused their analysis by either restricting the geographic area of analysis e.g. Bamdadinejad et al. (2021), or they target a specific coastline type e.g. Cheng et al. (2016). The constrained nature of these studies makes it difficult to understand the relative performance of approaches, and how well they may be able to characterise global coastlines in an automated and scalable fashion. Methodological development is also made difficult due to the absence of a globally distributed, labelled image dataset with which to train and benchmark different approaches. Of the few studies that have openly published labelled data, these are limited to a single geographic region as in Yang et al. (2020), or comprise images rendered from Google Earth at unknown locations and resolutions as in Li et al. (2018). In addition, reporting of model performance most often uses quantitative statistics and qualitative images of model performance on test sets that are drawn from the same imagery as the training set. There is limited description of how image level segmentation performance translates to geospatial accuracy of the defined land/water boundary, which may be of crucial importance for downstream, real-world use of model outputs or detailed description of how the performance of trained models generalises to other geographical regions. In summary, assessment of the relative performance of published methodologies is made difficult due to the current lack of openly available benchmark data, a lack of systematic methodological development and limited understanding of how image-level metrics relate to spatial quality.

Motivated by the opportunity to map global coastline morphology, this study aims to develop a CNN for detecting coastline morphology visible in Sentinel-2 satellite imagery, with an emphasis on geographic generalisability and the detection of small, linear coastal features. Unable to draw comparisons between pre-existing studies, we recognise the need for an open dataset designed for assessing coastline morphology detection methods. To encourage the consistent bench-marking of future coastline extraction techniques, we present and publish a new dataset for this purpose: the Sentinel-2 Water Edges Dataset (SWED). SWED is available as free, open data at https://openmldata.ukho.gov.uk. Using SWED, we present an end-to-end workflow for training and testing DL models. We take a systematic approach to optimising a CNN. Using a consistent U-Net-inspired model architecture, we implement models with standard loss functions to provide a base level of model performance. We then develop and apply geographically-weighted loss functions to focus model attention on the boundary between land and water. We therefore provide a systematic appraisal of the changes to model performance that is achievable through the development of novel boundary-focused loss functions. Finally, we assess the geographical generalisation of model performance, qualitatively assess model sensitivity to coastal features and examine the relationship between standard image level accuracy metrics and more geographically relevant metric.

## 2. Data and Methods

### 2.1. Sentinel-2 Water Edges Dataset

We introduce the Sentinel-2 Water Edges Dataset (SWED) for the development and bench-marking of coastline detection methods. SWED contains images captured by the European Space Agency's Sentinel-2 satellites between 2017 and 2021. When selecting images, we constrained our search to clear, cloud-free images by filtering the available catalog on the 'cloudy pixel percentage' metadata, and visually inspecting the returned results. No additional pre-processing to the source imagery was applied. The selected images are pre-allocated into train and test splits to support direct comparison of models. We annotated the selected images to create dense, pixel-level labels in two classes, 'water' and 'non-water'. The distribution of images was manually selected to ensure coverage of coastal environments that firstly span a wide range of geographies, as shown in Figure 1, and secondly a variety of coastline types at both high and low water conditions. During construction of the dataset, we were unable to identify an authoritative source of global coastline classification suitable for our needs, and therefore have created a list of coastline types compiled from our research (see Table 1). In addition, we included examples of narrow fine-grained features such as jetties and bridges. These features, whilst visible and recognisable to the human eye, are often overlooked by algorithms by virtue of their small size. The combination of images in SWED enables thorough testing of model sensitivity to small features, capability at recognising various coastline types and ability to geographically generalise. Here we explain the source data and annotation processes that created SWED.

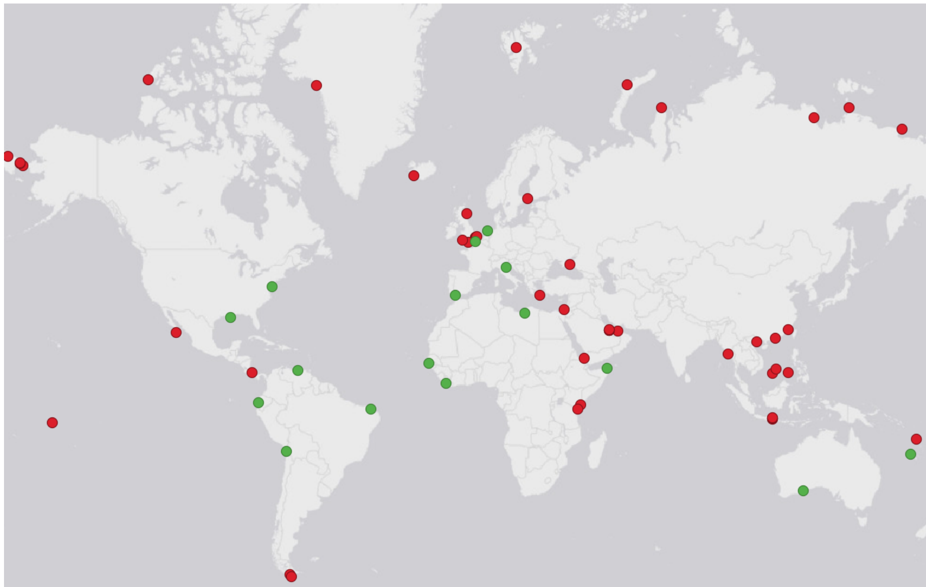

Figure 1: The distribution of labelled Sentinel-2 satellite imagery contained in the Sentinel-2 Water Edges Dataset (SWED). Annotations in two classes, water and non-water, were created for Sentinel-2 scenes at the highlighted coastal areas. 16 training and 49 testing locations are shown in green and red, respectively.

#### *2.1.1. Source Data*

The Sentinel-2 mission is a constellation of two Earth observation satellites equipped with multi-spectral imaging sensors, developed and operated by the European Space Agency's Copernicus Programme since 2015. The twin satellites, Sentinel-2A and 2B, systematically capture multi-spectral images over land and coastal waters with an approximately 5-day revisit period. We accessed Level 2A products (bottom-of-atmosphere reflectance) from the Sentinel-2 archive via the Copernicus Open Access Hub.

The highest spatial resolution of the Sentinel-2 Multi-Spectral Instrument (MSI) is 10 metres (see Table 2). For ease of use and compatibility with ma-

Table 1: Number of coastal types in the section of the Sentinel-2 Water Edges Dataset reserved for model testing, distributed by continent.

| Coastline type | Africa | Asia | Europe | North America | South America | Oceania | Total |
|---|---|---|---|---|---|---|---|
| Aquaculture | | 2 | | | | | 2 |
| Black Sand | | 2 | 8 | 4 | | | 14 |
| Boulders | | 12 | 10 | 2 | | | 24 |
| Breakwater | | 14 | 10 | | | | 24 |
| Bridge | | 2 | 8 | | | | 10 |
| Cliffs | | | 10 | | 2 | | 12 |
| Ice | | | 4 | 4 | | | 8 |
| Mangrove | 4 | 4 | | 4 | | 2 | 14 |
| Man-made | | 14 | 14 | | | 2 | 30 |
| Mud | 2 | 2 | 14 | 1 | | | 19 |
| Pebble or Shingle | 2 | 2 | 30 | 2 | 2 | 2 | 40 |
| Ramp | | 2 | 6 | | | | 8 |
| Rocky Shore | 2 | 12 | 18 | 4 | 4 | 4 | 44 |
| Salt Marsh | | | 14 | | | | 14 |
| Seawall | | 2 | 10 | | | | 12 |
| Weed | 2 | 8 | 9 | | 1 | | 20 |
| Wharf or Jetty | 2 | 10 | 16 | | | 2 | 30 |
| White Sand | 4 | 16 | 23 | 2 | 4 | 4 | 53 |

chine learning frameworks, any MSI bands at 20 or 60 metre resolution were re-sampled to 10 metre resolution using two-dimensional nearest neighbour interpolation. Consequently, one pixel within an image or array corresponds to a geographical plan area of 10m × 10m or $100m^2$.

Table 2: Sentinel-2 bands with wavelengths and spatial resolution. *Band 10 is only available with Level-1C products.

| Sentinel-2 MSI Band Descriptors | | | | |
|---|---|---|---|---|
| Band | Band Descriptor | S-2A Central Wavelength (nm) | S-2B Central Wavelength (nm) | Resolution (m) |
| Band 1 | Coastal Aerosol | 442.7 | 442.2 | 60 |
| Band 2 | Blue | 492.4 | 492.1 | 10 |
| Band 3 | Green | 559.8 | 559.0 | 10 |
| Band 4 | Red | 664.6 | 664.9 | 10 |
| Band 5 | Red Edge 1 | 704.1 | 703.8 | 20 |
| Band 6 | Red Edge 2 | 740.5 | 739.1 | 20 |
| Band 7 | Red Edge 3 | 782.8 | 779.7 | 20 |
| Band 8 | NIR | 832.8 | 832.9 | 10 |
| Band 8a | Red Edge 4 | 864.7 | 864.0 | 20 |
| Band 9 | Water Vapour | 945.1 | 943.2 | 60 |
| Band 10* | SWIR Cirrus | 1373.5 | 1376.9 | 60 |
| Band 11 | SWIR 1 | 1613.7 | 1610.4 | 20 |
| Band 12 | SWIR 2 | 2202.4 | 2185.7 | 20 |

### *2.1.2. Training Set Annotation Process*

We created annotations using a semi-supervised clustering approach using QGIS software (QGIS Development Team, 2021). Firstly, we rendered a

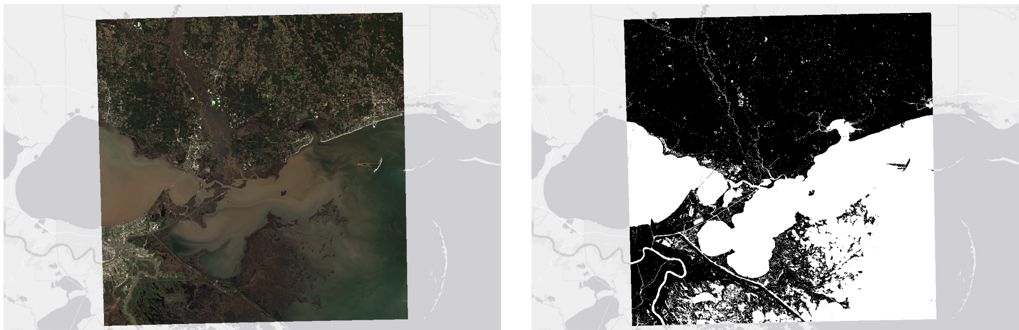

Figure 2: An example Sentinel-2 scene and corresponding segmentation mask taken from the Sentinel-2 Water Edges Dataset (SWED), showing pixels classified into 'non-water' and 'water' classes.

false colour image from spectral bands that display high contrast between water and non-water pixels. Sentinel-2 bands were selected for visualisation through trial and error to illicit the greatest visual contrast between water and non-water pixels. Whilst we did not find any band combination to be consistently successful across all training sites, the combinations of 8/11/4, 8/4/3 and 4/3/1 were found to be a good starting point when rendered in red, green and blue channels respectively (see Table 2 for further details of these bands). Secondly, we applied k-means clustering to the rendered image. The number of clusters $k$ was manually optimised to produce the best result for each individual image. Then, clusters were combined until two clusters remained, containing either water or non-water pixels. Lastly, we performed a visual comparison against high-resolution aerial imagery available in Google Earth and Bing Maps, and manually corrected any remaining mislabelled pixels to produce dense, pixel-level segmentation masks. An example image and mask are shown in Figure 2.

### 2.1.3. Test Set Annotation Process

A test set was created from a second batch of images that are geographically independent from those included in the training set. This is to create a test set that will test the ability of any coastline extraction methodology to generalise on imagery and geographies that are separate to that which was used for training. This is a more robust test structure than testing on a subset of the imagery that is used for training (López-Puigdollers et al., 2021) as it tests method's ability to generalise to variation in coastline, geography and Sentinel-2 scenes. The images in this set were curated to include a wide range of coastal types, varied tidal states and fine-grained coastal features, both natural and anthropogenic. Figure 1 shows the 49 selected geographic locations sampled in the test set. Two Sentinel-2 images at each location, one showing high water conditions, and one showing low water conditions, were labelled. Creation of target segmentation masks for this set involved careful, manual digitisation of the water/non-water boundary using the red, green, blue and near-infrared bands on a 256 x 256 pixel subset of each image by an experienced remote sensing analyst. The SWED test set therefore contains 98 image and segmentation mask pairs. The labelling effort applied to the test set was more intensive than the training set, to increase confidence in the overall position of the coastline within the test set labels.

## 2.2. Convolutional Neural Network Architecture

Convolutional Neural Networks consist of neuron layers that transform an image input into a desired output (e.g., an image classification label or segmented image mask). CNNs contain convolutional layers which apply trainable transformation filters in a moving window across an image, thus, learning to pick out features that are pertinent to the task they are being trained to perform. CNNs learn through a process of back propagation of error, defined by a loss function that compares model predictions to labelled data. The parameters that define model performance are updated via stochastic gradient descent (or a variant thereof) through repeated passes through training data (LeCun et al., 2015).

In order to provide a benchmark of CNN model performance against the SWED test set, we trained four CNN models with identical model architectures using four different loss functions. Our deep, convolutional neural network design is based on U-Net, a CNN architecture developed by Ronneberger et al. (2015) for the segmentation of biomedical images. Since publication, this architecture has proved capable of generalising to many semantic segmentation tasks (Galeone, 2019), including satellite image segmentation and detection of the coastline (Shamsolmoali et al., 2019; Li et al., 2018; Chu et al., 2019; Yang et al., 2020). The architecture is structured in an encoder-decoder pattern, with skip connections concatenating feature information extracted in the encoder path with information in the decoder path. Our implementation is shown diagrammatically in Figure 3.

The encoder path is composed of four blocks. Each block contains two convolutional layers, each with a 3 x 3 kernel and Exponential Linear Unit (ELU) activation (Clevert et al., 2016). A batch normalisation layer follows each convolutional layer. Each block finishes with a max-pooling layer with a 2 x 2 pool size. The decoder path has four corresponding blocks. Each decoder block contains an up-sampling layer of size 2 x 2, followed by two convolutional layers, again using a 3 x 3 kernel, ELU activation, and batch

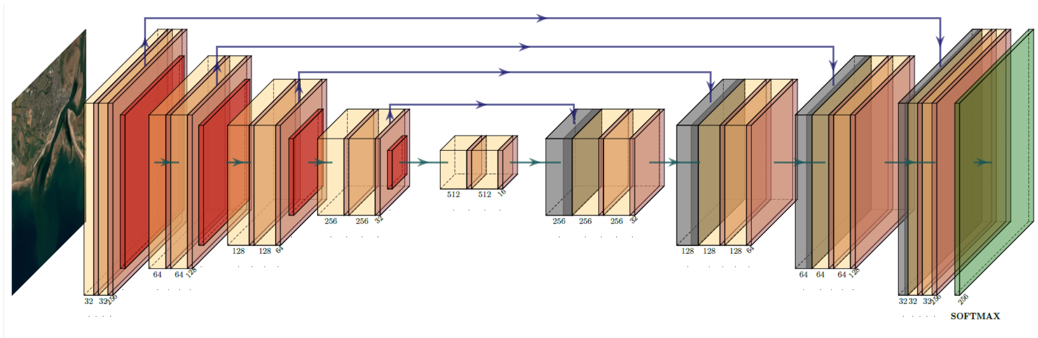


Figure 3: A schematic representation of the U-Net-based convolutional neural network architecture used to segment Sentinel-2 imagery into water and non-water classes. The same architecture was optimised multiple times with different loss functions, to compare the sensitivity of the resulting models to coastal features. Convolutional, batch normalisation, max-pooling, up-sampling and softmax layers are shown in yellow, brown, red, grey and green respectively. Arrows denote skip connections.

normalisation layer. The last layer is a convolutional layer with 2 filters of size 1 x 1 with softmax activation.

### *2.3. Loss Functions*

#### *2.3.1. Standard Loss Functions*

A U-Net model optimised using cross-entropy loss provides an initial benchmark for segmentation performance on the SWED dataset. This model represents the 'default' for coastline detection performance without any task-specific adaptations. Cross-entropy loss examines each pixel within an image individually, to compare the predicted class to the target class. Let $p$ indicate probability $P$ that the pixel is of the positive class label $Y$ for a pixel with a 0 or 1:

$$P(Y = 0) = p \quad \text{and} \quad P(Y = 1) = 1 - p \tag{1}$$

Where $\hat{p}$ is the predicted probability an observation is of the positive class, cross-entropy loss is given as:

$$CE(p, \hat{p}) = -(p \log(\hat{p}) + (1 - p) \log(1 - \hat{p})) \tag{2}$$

Each pixel has equal weighting, with the error at each pixel calculated discretely. Use with imbalanced training data may limit the performance of a model, as it may struggle to learn the smaller class due to the dominance of the more frequent class on the loss value (Lin et al., 2017).

Sørensen–Dice loss (SDL) was selected for comparison as it is recommended for image segmentation tasks when a class imbalance is present in the training data (Sudre et al., 2017). Based on a reformulation of the Sørensen–Dice coefficient (SDC) as a loss function, first proposed by Milletari et al. (2016), it measures overlap between predicted and target classes, aiming to assess the quality of segmentations rather than the pixel-wise accuracy. For a prediction with true positive (TP), false negative (FN) and false positive (FP) results, SDC is defined as follows:

$$SDC = \frac{2TP}{2TP + FN + FP} \tag{3}$$

Formulated as loss function, SDL is as follows:

$$SDL(p, \hat{p}) = 1 - \frac{2 \sum_{i,j} (p \odot \hat{p})_{i,j} + \epsilon}{\sum_{i,j} p_{i,j} + \sum_{i,j} \hat{p}_{i,j} + \epsilon} \tag{4}$$

The numerator of SDL is approximated by summing all the values from the element-wise multiplication between $p$ and $\hat{p}$, and the denominator is obtained by summing all the elements $p$ and $\hat{p}$. A small value $\epsilon$ is added to the denominator to avoid division by zero, and added to the numerator to smooth the result (Planche and Andres, 2019). SDL, as with cross-entropy, is calculated across all pixels, without any geographic or feature-specific focus.

*2.3.2. Edge-weighted Loss Functions*

Initial experiments suggested that the error-averaging intrinsic to cross-entropy and SDL limits learning potential if the segmentation task contains complex small or linear features at segment boundaries (as seen in coastline boundaries within satellite images). The selection of a suitable loss function for coastline segmentation is imperative in driving the training process to find optimal parameters as it defines, in part, the parameter adjustments (Galeone, 2019). In response, we propose two custom loss functions designed to focus network optimisation on error at the segmentation boundary.

Firstly, we trialled a novel loss function termed 'Sobel-edge loss'. This approach leverages the Sobel edge detection algorithm (Vincent and Folorunso, 2009) to extract the edges between segments. Performed on both predicted and target segmentation, a loss is obtained by comparing the two sets of edges. Our hope is that training will drive the predicted segments to have increasingly similar edges to the target segments. We define Sobel-edge loss as the Mean Square Error between edges detectable on targets and those detectable on predictions, through the application of Sobel edge detection filters. A Sobel edge detector filter uses two kernels that apply a convolution operation to an input (Vincent and Folorunso, 2009). For an image $A$, the two filters compute gradients along the x and y axes as below:

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} * A \tag{5}$$

and

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * A \tag{6}$$

Typically, during Sobel edge detection the magnitude of the gradients $G$ is then calculated as:

$$G = \sqrt{G_x^2 + G_y^2} \tag{7}$$

Formulated as a loss function, where $i$ is a training image sample, the Sobel loss between the target $p$ and the prediction $\hat{p}$ is therefore:

$$\text{Sobel}(p, \hat{p}) = \frac{1}{n} \sum_{i=1}^{n} \Big( G_{p_i} - G_{\hat{p}_i} \Big)^2 \tag{8}$$

The second custom loss function tested applies a modification to SDL. Using the magnitude $G$ of the Sobel edges extracted from the target $p$ as a weight matrix, the predictions $\hat{p}$ are weighted along segment boundaries, amplifying the contribution to the loss from the boundary area. With a weight matrix $W = G_p$, we therefore define the Weighted Sørensen–Dice loss (W-SDL) as:

$$\text{W-SDL}(p, \hat{p}) = 1 - \frac{2 \sum_{i,j} (p \odot W \hat{p})_{i,j} + \epsilon}{\sum_{i,j} p_{i,j} + \sum_{i,j} W \hat{p}_{i,j} + \epsilon} \tag{9}$$

### 2.3.3. Training Phase

Four models were created, each using a different loss function but otherwise identical in every aspect. Two of the models were optimised using standard loss functions: Categorical cross-entropy loss and SDL. Two further models were trained using Sobel-edge loss and W-SDL. The models were trained on the SWED training set, using 23807 and 2661 256x256 pixel patches for training and validation sets respectively. Training data was randomly shuffled before input. Each model variant was trained for 50 epochs, with early stopping if the validation loss did not improve after 10 epochs. The learning rate was reduced by a factor of 0.1 when the validation loss plateaued for more than 5 epochs. At the end of each epoch, the model weights were saved if the validation loss improved, resulting in the final model weights being those associated with the smallest validation loss over the course of training.

All models were implemented in TensorFlow 2.2 using Python 3.8 and model training was completed using an Amazon Web Services EC2 instance with 8 NVIDIA V100 Tensor Core GPUs.

## 2.4. Model Assessment Methodology

Model performance was evaluated in four ways:

1. Quantitative evaluation using standard metrics for image segmentation. These are accuracy, balanced accuracy, Precision, Recall, Cohen's Kappa, F1-score, Jaccard Index and Matthew's Correlation Coefficient (MCC).
2. The ability of the model to geographically generalise across the six continents and different tidal states sampled within the SWED test

dataset. Each test location was cross-referenced with the World Bank Land Boundaries dataset to define continent attributes (World Bank, 2020).

3. Qualitative assessment of model sensitivity to fine-grained coastal features. Predictions on all 98 test images were visually assessed and patterns of consistent misclassification by the models were noted as well as the presence/absence of small-scale fine coastal detail at the segmentation boundary.

4. Positional accuracy of the defined water/non-water boundary.

### *2.4.1. Quantitative evaluation metrics*

Quantitative evaluation metrics were calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$\text{Balanced Accuracy} = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{11}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{13}$$

$$\text{F1-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{15}$$

$$\text{Jaccard} = \frac{TP}{TP + FP + FN} \tag{16}$$

where $TP$ denotes true positives - correctly predicted water pixels. $TN$ denotes true negatives - correctly predicted non-water pixels. $FP$ denotes false positives - non-water pixels predicted to be water. $FN$ denotes false negatives - water pixels predicted to be non-water.

We also report Cohen's kappa score . Where $p_o$ is the observed accuracy and $p_e$ is the expected accuracy given random chance:

$$\text{Cohen's kappa} = \frac{(p_o - p_e)}{(1 - p_e)} \tag{17}$$

*2.4.2. Positional accuracy*

We determined the positional accuracy of predicted coastlines (the boundary between water and non-water pixels) using an adaptation of the "Trimap" method described by Kohli et al. (2008). The original Trimap method computes performance metrics within buffered regions around the target object boundary. We developed and used a variant to the original method that calculates the percentage of predicted boundary that falls within buffered regions of the target object boundary. Calculation of percentages within a tolerance allows us to describe results using statements of accuracy, such as "$x$% of the predicted coastline is within $y$ metres of the target boundary".

To convert segmented images to a predicted coastline, we used an implementation of the marching squares algorithm (Lorensen and Cline, 1987) available in the scikit-image Python library (van der Walt et al., 2014). Values between adjacent pixels were linearly interpolated and a line was drawn along a contour of constant value. In this work, a segmented image consists of two classes, water and non-water, represented by pixel values of either 0 or 1 and therefore the 0.5-valued contour was taken to be the coastline.

## 3. Results

Table 3: Quantitative evaluation of model performance for four models under test, identified by the loss function used during training, e.g. CXE: Categorical Cross-entropy Loss, SDL: Sørensen–Dice Loss, W-SDL, Weighted Sørensen–Dice Loss, Sobel: Sobel-edge Loss. Best results for each metric are emboldened.

| | Metrics | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Loss function | Accuracy | Balanced Accuracy | Precision | Recall | Cohen's Kappa | F1-score | Jaccard Index | MCC |
| CXE | **0.937** | **0.910** | **0.916** | 0.948 | **0.820** | **0.922** | **0.875** | **0.835** |
| SDL | 0.905 | 0.876 | 0.842 | **0.987** | 0.740 | 0.892 | 0.834 | 0.769 |
| W-SDL | 0.886 | 0.868 | 0.869 | 0.887 | 0.711 | 0.860 | 0.784 | 0.731 |
| Sobel | 0.934 | 0.909 | 0.910 | 0.948 | 0.817 | 0.917 | 0.871 | 0.834 |

### 3.1. Quantitative metrics

Performance metrics, computed on the SWED test set for each model, are presented in Table 3. The model trained using Categorical cross-entropy loss ("CXE model") achieved the best performance across all metrics apart from recall. The model using Sobel-edge loss ("Sobel-edge model") achieves almost equivalent performance by these measures. The model using a Weighted

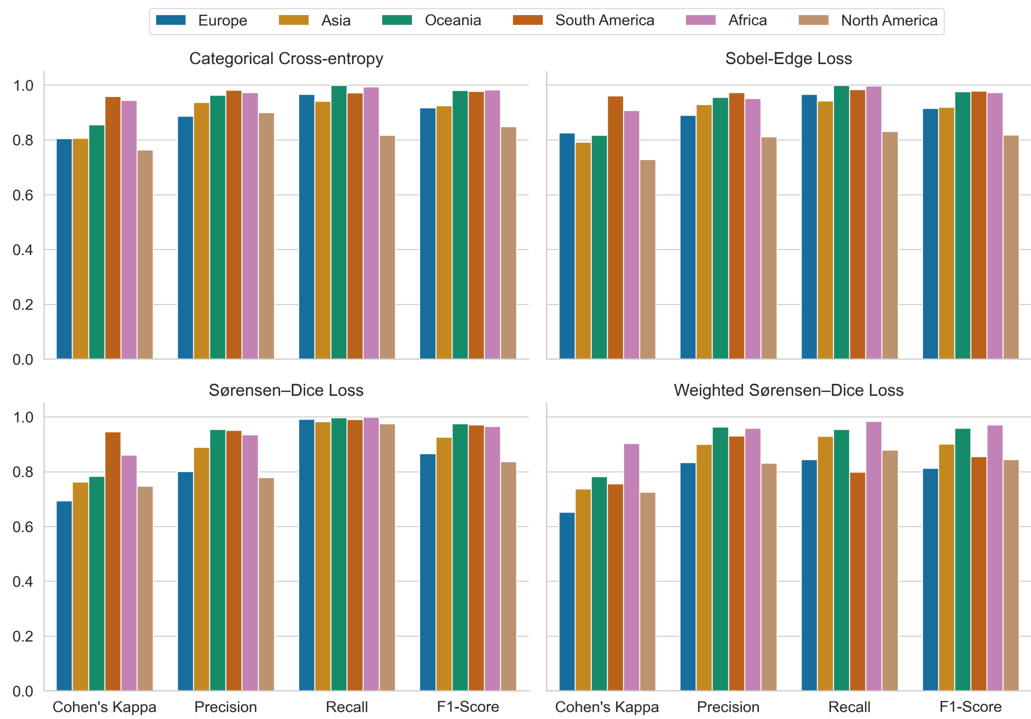


Figure 4: Performance metrics Cohen's Kappa score, Precision, Recall and Mean F1-score for four models created using different loss functions, by geographic region.

Sørensen–Dice loss ("W-SDL model") was the worst performer across all metrics except precision.

### *3.2. Generalisability to geographic region and tidal state*

Figure 4 examines the geographical variation of Cohen's Kappa, precision, recall and F1-score for each of the tested models. There is variety in model performance across the continents from which the SWED test set is sampled. For the F1-score statistic, a similar performance is recorded at continental scale for the CXE and Sobel-edge models, with the model trained using Sørensen–Dice loss ("SDL Model") exhibiting a similar behaviour apart from the European performance, which is reduced. Unlike the other three models, the W-SDL model has a reduced performance in South America in comparison to the competing models. For Cohen's Kappa score, most models perform best in African and South American continents. The W-SDL model is the exception, once again the worst performer in South America. South America, Oceania, Asia and Africa have the highest precision and recall across models, with Europe and North America tending to have a lower precision and recall across all models.

Figure 5 illustrates the performance of each model on images showing either high or low tide conditions. There is a consistent trend of slightly improved performance on high tide images, this is seen across all model results.

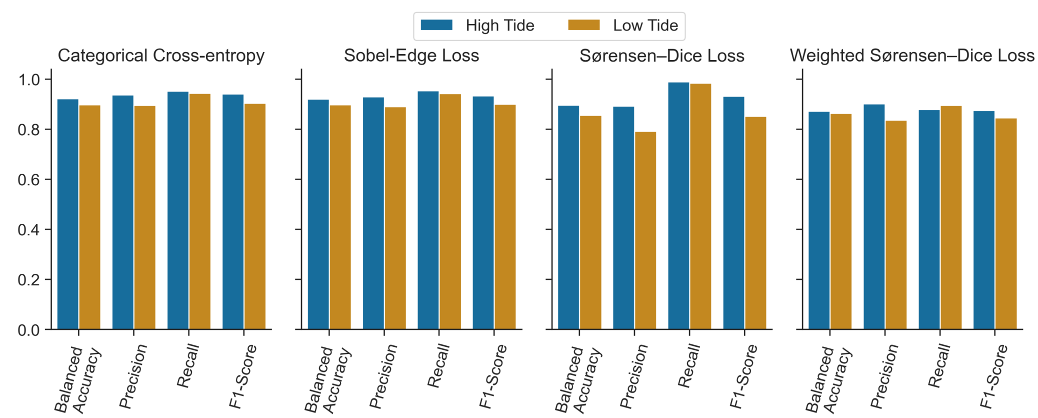


Figure 5: Quantitative metrics for four models created using different loss functions, when shown unseen images depicting conditions at either low or high tide.

### 3.3. Sensitivity to small-scale geographic detail

Qualitative analysis demonstrated the CXE and SDL models were less sensitive to small or fine-grained features in comparison to those trained using edge-weighted loss functions.

Examples from the test set of SWED containing instances of narrow, linear coastal features are shown in Figure 6. Qualitative assessment was performed on all 98 test label/image pairs, but for clarity of description, here we present examples that allow us to describe the differences in the performance of the different models. In Row A, a linear anthropogenic feature is interpreted clearly by the human eye. All models apart from the SDL model detect this feature, however, the Sobel-edge model defines greater detail while the W-SDL model contains a large number of false negative predictions (i.e water predicted as non-water). In Row B, a sand barrier is detected by all four models with the Sobel-edge model detecting a greater amount of fine detail. Row C contains a natural feature of a sand barrier and tidal sandy islands. The SDL model fails to detect all these features, whilst the CXE model fails to detect the sandy islands. The two models trained with edge-weighted loss functions (Sobel-edge and W-SDL) have superior performance on this test image. Row D features three narrow anthropogenic features that protrude from the land into the sea. These are narrow features in width that approach the best native resolution of the Sentinel-2 satellite of 10m/one pixel. Whilst the CXE and SDL models detect the wider parts of the features, the linear portion is undetected. Sobel-edge and W-SDL models detect the linear and wide base parts of these features, with the W-SDL model having superior performance. In Row E there are natural and anthropogenic features that

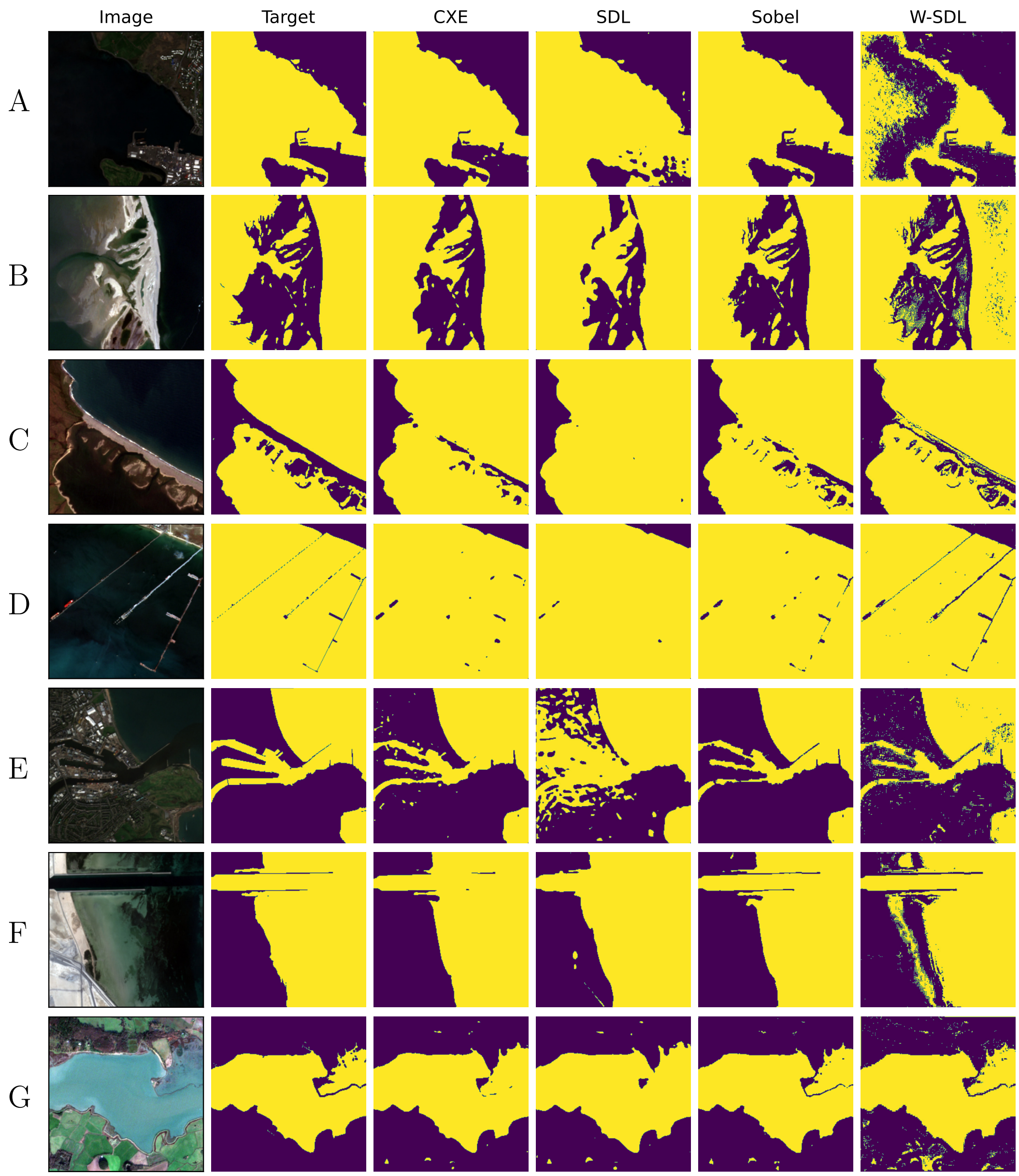


Figure 6: Visualised segmentation results. The first and second columns depict the source image and the target label respectively. The subsequent columns each depict the predicted segmentation map from a different model, identified by the loss function used during training, e.g., CXE: Categorical Cross-entropy Loss, SDL: Sørensen–Dice Loss, Sobel: Sobel-Edge Loss and W-SDL: Weighted-Sørensen–Dice Loss.

protrude into the water, and from the water into the land. The Sobel-edge model demonstrates superior detection of the narrow linear features, with the CXE and SDL models lacking detail. In Rows F and G, the Sobel-edge model's superior performance for the detection of small narrow features is demonstrated in multiple locations, with multiple types of natural and anthropogenic features.

### *3.4. Positional accuracy of the defined coastline boundary*

Figure 7 compares the positional accuracy of the coastlines defined from the test predictions made by each of the four models. The Sobel-edge model has consistently the greatest proportion of predicted coastlines within buffer radii, whilst the CXE model demonstrates a similar performance that is superior in comparison to the SDL and W-SDL models. Nearly 60% of the predicted coastline for the Sobel and CXE models is within a buffer radius of 20m (or two pixel widths) of the target coastline, whilst for W-SDL this drops to $< 30\%$. At a buffer radius of 50m, nearly 75% of the Sobel and CXE model coastlines are within the buffer and at a radius of 90m this increases to just under 80%. In contrast W-SDL only reaches $> 45\%$ within a buffer of 90m.

## 4. Discussion

In this study we have described the Sentinel-2 Water Edges Dataset, which we have developed and published to enable the comparison of future developments of automated coastline extraction techniques. We have benchmarked the performance of four deep learning models for the definition of coastline morphology and assessed model performance using qualitative and

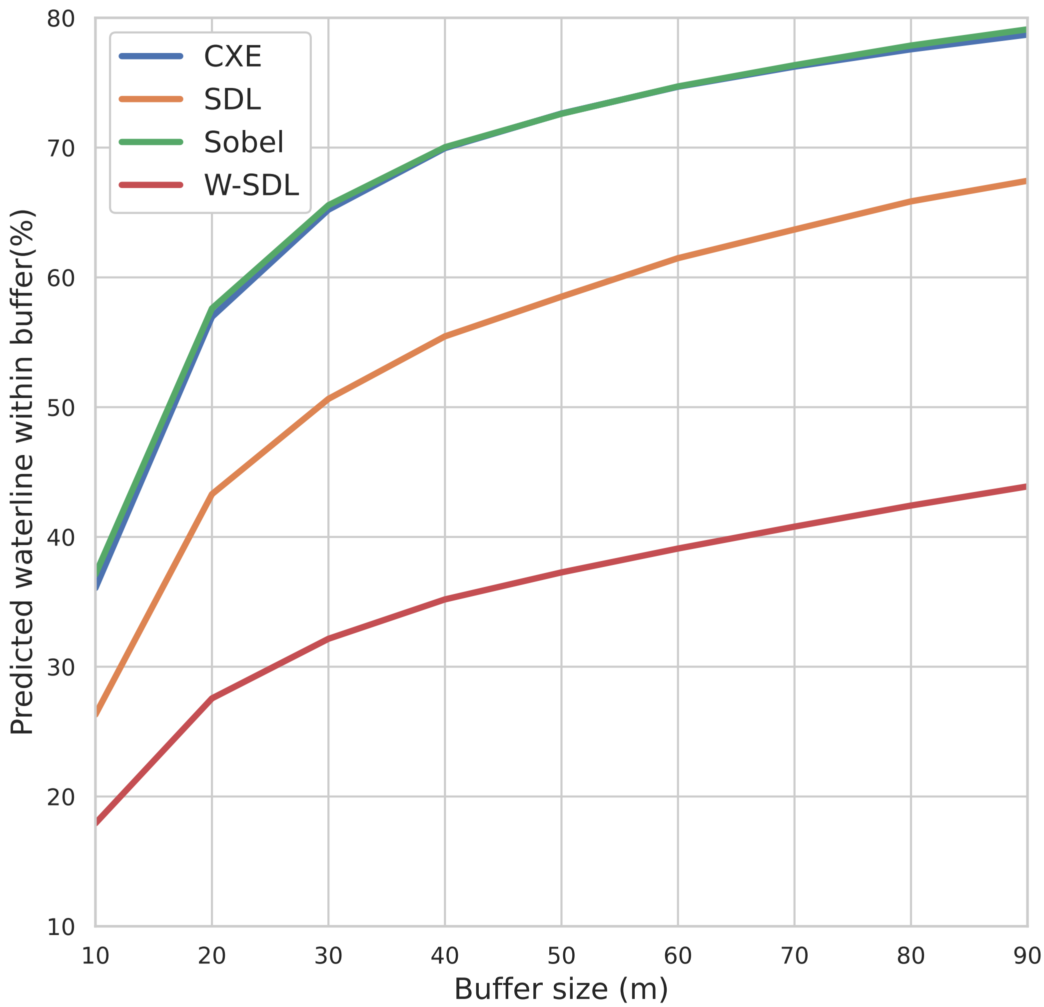


Figure 7: Comparison of what percentage of coastline generated by four different models under test is contained within a buffered region around the target coastal boundary. An increase in the y-axis is interpreted as producing coastlines with greater positional accuracy. The legend identifies the loss function used to train the model, e.g., CXE: Categorical Cross-entropy Loss, Sobel: Sobel-edge Loss, SDL: Sørensen–Dice Loss, W-SDL: Weighted Sørensen–Dice Loss.

quantitative analysis. Our analyses demonstrate that it is not possible to identify the important qualitative differences in model performance detailed in Section 3.3 through the comparison of models using the quantitative statistics used in Section 3.1. This is an important finding as it demonstrates the insensitivity of current image segmentation performance metrics to coastal features such as pipelines, piers and bridges, thus demonstrating the need for qualitative assessment of model performance.

The qualitative analysis presented in Section 3.3 demonstrates that the choice of loss function affects model performance and sensitivity of final trained models to small coastal features. Models with similar quantitative test performance are shown to possess different sensitivities to small scale coastal detail, thus, future researchers should consider a full range of quantitative and qualitative performance assessment methodologies when designing and testing coastline extraction methodologies. The Sobel-edge loss proposed by this study produced the best performing model by qualitative but not quantitative analysis. Sobel-edge loss was particularly effective at the detection of narrow coastline features, such as bridges, breakwaters and jetties. These small features are disproportionately important for their size, occurring in ports and developing areas of human influence, but were often missed by standard image segmentation loss functions. Sobel-edge loss was shown to promote their persistence through the deep network. It was, however, difficult to quantify the ability of models to maintain fine detail. Such features may only be a few pixels in size and consequently their presence or absence did not greatly effect segmentation performance metrics or assessments of coastline positional accuracy. To fully understand the nature of predictions, visualisation and a hybrid qualitative and quantitative assessment was required. Relying on quantitative statistics alone would result in the choice of a model that was insensitive to such fine coastal features.

The novel loss function presented here, Sobel-edge Loss, demonstrates a similar quantitative performance to the best performing model. However, qualitative assessment of model performance demonstrated the superior performance of the Sobel-edge model in the detection of narrow, detailed coastal features. Therefore, we would recommend the use of the Sobel-edge loss function for the training of future coastline detection CNN models.

Creating a single ML or DL model able to geographically generalise across large regions is a challenge (Wang et al., 2017), that requires specific attention in the design of model testing strategies (Waldner and Diakogiannis, 2020; López-Puigdollers et al., 2021). In this study we compared model test metrics at the continental scale to define a baseline of geographic variation in model performance. Our analysis demonstrates that the performance of all four tested models varies across the different continents and that there are similarities between the CXE and Sobel-edge models. At this stage we are unable to describe whether this variation in test performance is a result of model sensitivity to coastal features, other landscape features within SWED test imagery or some other phenomena that may lead to variation in test imagery e.g. atmospheric interference as a result of aerosols or particulates. Indeed, we have compared performance at a continental scale, but it could be that a more fine-grained analysis at the country or landscape level may provide further insight into variation in model performance. Whilst care was taken to ensure that both training and test data contained examples of a diverse range of coastline types, we were unable to evaluate model performance by coastline type in a quantitative manner. For this to be possible, pixel-level labels of coastline type are required, but in this study only image-level labels describing coastline types were prepared. An individual image very often contained more than one type of coastline, e.g. an image may show a beach, a rocky shore and a pier. Consequently, we were able to describe the total count of examples depicting various coastline types, as detailed in Table 1, but not which pixels within an image belong in which class. The creation of pixel-level coastline class labels is therefore a recommendation for improving SWED, as it would allow for detailed evaluation results with respect to coastline type. In lieu of an understanding of what is the optimal geographical scale to understand variation in model performance, continental scale was chosen as a pragmatic approach - as it was possible to cross reference our test image locations with readily available data. Improving understanding of the difficulties of generalising model performance geographically is recommended, so that future models may be more robustly tested and thus deployed with confidence to areas outside of original model training. We also compared model test metrics at different tidal states. The results indicated that performance was greater on high-tide images than on low-tide images of the same areas. An explanation may be the differing characteristics of a low-tide line versus its high-tide counterpart. A low-tide line may have an ambiguous boundary with no meaningful shape to it (Vos et al., 2019). In a low-tide image, the landward side may be wet, for example saturated mud or sand, providing reduced contrast between the two water/not water classes and therefore be more difficult for classifiers to distinguish (Ryu et al., 2002;

Bishop-Taylor et al., 2021). When the tide is high, the coastline is more likely to be formally defined by a cliff, wall or other structure (Vos et al., 2019), and a CNN can leverage spatial information relating to coastline shape to make a prediction. Therefore, we recommend future training data to include imagery at different tidal states, with comprehensive examples at low and mid-tides, in addition to showing diverse examples of coastline morphology.

A limitation of this study is the absence of in-situ measurements of coastal position for creating training data and evaluating results which could offer increased precision and confidence in labels. However, it is impractical to take in-situ measurements for a study with global scope, and we were subsequently constrained to visual interpretation of coastline position from imagery as a method of creating training and testing data. Variations in visual interpretation may introduce uncertainty that could make it problematic to assess performance. The problem of imperfect labels is unresolved and was occasionally apparent in this study in cases where predictions appeared to be more accurate than labels suggested. However, the visual interpretation of remote sensing imagery is common in studies using image analysis and machine learning techniques with remote sensing imagery (Bunting et al., 2018; Cheng et al., 2016), as it remains the only practical way of creating large, geographically distributed training and testing sets. Another limitation may be a lack of variation in training examples, which were constrained to clear, cloud-free images and are therefore not representative of all Sentinel-2 images possible. As a result, segmentation models trained using these examples will likely produce poor predictions if deployed indiscriminately on Sentinel-2 imagery e.g., on images that are partially obstructed by cloud.

Putting these limitations aside, the publication of SWED and the methodological advancements described here are important steps towards the ambitious goal of creating a coastline detection system that can be scaled to the global (coastal) catalogue of Sentinel-2 imagery. With care, models trained using the methods described here, with the SWED dataset, could be used within an appropriate data pipeline to define the world's coastal morphology, at a hitherto unavailable spatial and temporal resolution (e.g. 10m resolution of Sentinel-2 imagery with a 5 day revisit time). What's more, given the availability of past and future imagery, once a baseline coastal morphology dataset is available, repeated periodic monitoring should be possible thus allowing for the application of change detection techniques. In order to achieve this aim, further work is required to determine the accuracy of trained CNN models in imagery over time, improving the generalisability of models across coastline types at different tidal states and of the identification of changes to natural and anthropogenic features.

## 5. Conclusions

In this paper we introduce the Sentinel-2 Water Edges Dataset and the Sobel-edge and Weighted Sørensen-Dice loss functions. These loss functions were developed with the specific aim of targeting model training attention to fine-scale image detail at the boundary of segmentation targets. We demonstrate the superior performance of a baseline U-Net model optimised using the Sobel-edge Loss in comparison to more commonly used loss functions. Test results on the SWED dataset illustrate the improvement in performance through the use of Sobel-edge loss with fine-grained coastal detail detected in test images. SWED and the Sobel-edge Loss may now be used to optimise image segmentation networks for coastline detection.

## 6. Data Access

The Sentinel-2 Water Edges Dataset can be obtained by visiting openml-data.ukho.gov.uk and used under the Geospatial Commission Data Exploration license.

## 7. Acknowledgements

This work was supported by the UK Hydrographic Office. The authors would like to thank the entire Data Science and Remote Sensing teams at the UK Hydrographic Office for their support, in particular David Stephens and Pascal Philipp for their valuable discussions and subject matter expertise.

## References

Bamdadinejad, M., Ketabdari, M.J., Chavooshi, S.M.H., 2021. Shoreline extraction using image processing of satellite imageries. Journal of the Indian Society of Remote Sensing 49, 1–11.

Bishop-Taylor, R., Nanson, R., Sagar, S., Lymburner, L., 2021. Mapping Australia's dynamic coastline at mean sea level using three decades of Landsat imagery. Remote Sensing of Environment 267, 112734.

Bishop-Taylor, R., Sagar, S., Lymburner, L., Beaman, R.J., 2019. Between the tides: Modelling the elevation of Australia's exposed intertidal zone at continental scale. Estuarine, Coastal and Shelf Science 223, 115–128.

Bunting, P., Rosenqvist, A., Lucas, R.M., Rebelo, L.M., Hilarides, L., Thomas, N., Hardy, A., Itoh, T., Shimada, M., Finlayson, C.M., 2018. The global mangrove watch — a new 2010 global baseline of mangrove extent. Remote Sensing 10, 1669.

Burningham, H., French, J., 2017. Understanding coastal change using shoreline trend analysis supported by cluster-based segmentation. Geomorphology 282, 131–149.

Cao, W., Zhou, Y., Li, R., Li, X., 2020. Mapping changes in coastlines and tidal flats in developing islands using the full time series of Landsat images. Remote Sensing of Environment 239, 111665.

Cheng, D., Meng, G., Cheng, G., Pan, C., 2016. SeNet: Structured edge network for sea–land segmentation. IEEE Geoscience and Remote Sensing Letters 14, 247–251.

Chu, Z., Tian, T., Feng, R., Wang, L., 2019. Sea-land Segmentation with Res-UNet and fully connected CRF, in: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, pp. 3840–3843.

Clevert, D., Unterthiner, T., Hochreiter, S., 2016. Fast and accurate deep network learning by Exponential Linear Units (ELUs), in: Bengio, Y., LeCun, Y. (Eds.), Proceedings of the 4th International Conference on Learning Representations.

Edmonds, D., Caldwell, R., Brondizio, E., Siani, S., 2020. Coastal flooding will disproportionately impact people on river deltas. Nature Communications 11, 4741.

European Commission, 2013. Legal notice on the use of Copernicus Sentinel Data and Service Information. https://sentinel.esa.int/documents/247904/690755/Sentinel_Data_Legal_Notice. Accessed 2021-03-17.

Galeone, P., 2019. Hands-On Neural Networks with TensorFlow 2.0. Packt Publishing.

Heene, G., Gautama, S., 2000. Optimisation of a coastline extraction algorithm for object-oriented matching of multisensor satellite imagery, in: IEEE International Geoscience and Remote Sensing Symposium, pp. 2632–2634.

Karantzalos, K.G., Argialas, D., Georgopoulos, A., 2002. Towards automatic detection of coastlines from satellite imagery, in: 14th International Conference on Digital Signal Processing Proceedings, pp. 897–900.

Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on convolutional neural networks (CNN) in vegetation remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 173, 24–49.

Klinger, T., Ziems, M., Heipke, C., Hannover Schenke, H.W., Ott, H., 2012. Automated extraction of the Antarctic coastline using snakes. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 38.

Kohli, P., Ladicky, L., Torr, P.H.S., 2008. Robust higher order potentials for enforcing label consistency, in: Preceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8.

Kuschnerus, M., Lindenbergh, R., Vos, S., 2021. Coastal change patterns from time series clustering of permanent laser scan data. Earth Surface Dynamics 9, 89–103.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Li, R., Liu, W., Yang, L., Sun, S., Hu, W., Zhang, F., Li, W., 2018. DeepUNet: A deep fully convolutional network for pixel-level sea-land segmentation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11, 3954–3962.

Li, W., Gong, P., 2016. Continuous monitoring of coastline dynamics in western Florida with a 30-year time series of Landsat imagery. Remote Sensing of Environment 179, 196 – 209.

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection, in: Proceedings of the 2017 IEEE International Conference on Computer Vision, pp. 2999–3007.

Liu, H., Jezek, K.C., 2004. Automated extraction of coastline from satellite imagery by integrating canny edge detection and locally adaptive thresholding methods. International Journal of Remote Sensing 25, 937–958.

López-Puigdollers, D., Mateo-García, G., Gómez-Chova, L., 2021. Benchmarking deep learning models for cloud detection in Landsat-8 and Sentinel-2 Images. Remote Sensing 13, 992.

Lorensen, W.E., Cline, H.E., 1987. Marching cubes: A high resolution 3d surface construction algorithm, in: Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques, Association for Computing Machinery. pp. 163—169.

Loveland, T.R., Dwyer, J.L., 2012. Landsat: Building a strong future. Remote Sensing of Environment 122, 22–29.

Luijendijk, A., Hagenaars, G., Ranasinghe, R., Baart, F., Donchyts, G., Aarninkhof, S., 2018. The state of the world's beaches. Scientific Reports 8, 6641.

Martinez, M., Intralawan, A., Vázquez, G., Pérez-Maqueo, O., Sutton, P., 2007. The coasts of our world: Ecological, economic and social importance. Ecological Economics 63, 254–272.

Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-Net: Fully convolutional neural networks for volumetric medical image segmentation, in: Proceedings of the 2016 Fourth International Conference on 3D Vision, pp. 565–571.

Murray, N., Phinn, S., DeWitt, M., Ferrari, R., Johnston, R., Lyons, M., Clinton, N., Thau, D., Fuller, R.A., 2018. The global distribution and trajectory of tidal flats. Nature 565, 222–225.

Oehmcke, S., Thrysøe, C., Borgstad, A., Salles, M.A.V., Brandt, M., Gieseke, F., 2019. Detecting hardly visible roads in low-resolution satellite time series data, in: IEEE International Conference on Big Data (Big Data), pp. 2403–2412.

Paravolidakis, V., Ragia, L., Moirogiorgou, K., Zervakis, M., 2018. Automatic coastline extraction using edge detection and optimization procedures. Geosciences 8, 407.

Parente, L., Taquary, E., Silva, A.P., Souza, C., Ferreira, L., 2019. Next generation mapping: Combining deep learning, cloud computing, and big remote sensing data. Remote Sensing 11, 2881.

Planche, B., Andres, E., 2019. Hands-On Computer Vision with TensorFlow 2. Packt Publishing Ltd.

QGIS Development Team, 2021. QGIS Geographic Information System. QGIS Association. Accessed 2020-08-11.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Medical Image Computing and Computer-Assisted Intervention, Springer International Publishing. pp. 234–241.

Ryu, J.H., Won, J.S., Min, K.D., 2002. Waterline extraction from Landsat TM data in a tidal flat: A case study in Gomso Bay, Korea. Remote Sensing of Environment 83, 442–456.

Shamsolmoali, P., Zareapoor, M., Wang, R., Zhou, H., Yang, J., 2019. A novel deep structure U-Net for sea-land segmentation in remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12, 3219–3232.

Sudre, C., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J., 2017. Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations, in: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp. 240–248.

Sultana, F., Sufian, A., Dutta, P., 2020. Evolution of image segmentation using deep convolutional neural network: A survey. Knowledge-Based Systems 201-202, 106062.

Toure, S., Diop, O., Kpalma, K., Maiga, A.S., 2019. Shoreline detection using optical remote sensing: A review. ISPRS International Journal of Geo-Information 8.

Uddin, K., Khanal, N., Chaudhary, S., Maharjan, S., Thapa, R.B., 2020. Coastal morphological changes: Assessing long-term ecological transformations across the northern Bay of Bengal. Environmental Challenges 1, 100001.

Vincent, O.R., Folorunso, O., 2009. A descriptive algorithm for Sobel image edge detection, in: Proceedings of the 2009 Informing Science IT Education Conference.

Vos, K., Harley, M.D., Splinter, K.D., A, S.J., Turner, I.L., 2019. Sub-annual to multi-decadal shoreline variability from publicly available satellite imagery. Coastal Engineering 150, 160 – 174.

Waldner, F., Diakogiannis, F.I., 2020. Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. Remote Sensing of Environment 245, 111741.

van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T., and the Scikit-image contributors, 2014. Scikit-image: image processing in Python. PeerJ 2, 453.

Wang, R., Camilo, J., Collins, L.M., Bradbury, K., Malof, J.M., 2017. The poor generalization of deep convolutional networks to aerial imagery from new geographic locations: an empirical study with solar array detection, in: Proceedings of the 2017 IEEE Applied Imagery Pattern Recognition Workshop, pp. 1–8.

Wong, P., Losada, I., Gattuso, J., Hinkel, J., Khattabi, A., McInnes, K., Saito, Y., Sallenger, A., 2014. Coastal systems and low-lying areas, in: Field, C., Barros, V., Dokken, D., Mach, K., Mastrandrea, M., Bilir, T., Chatterjee, M., Ebi, K., Estrada, Y., Genova, R., Girma, B., Kissel, E., Levy, A., MacCracken, S., Mastrandrea, P., White, L. (Eds.), Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 361–409.

World Bank, 2020. World Bank Official Boundaries. `https://datacatalog.worldbank.org/search/dataset/0038272/World-Bank-Official-Boundaries`. Accessed 17-09-2020.

Wyles, K.J., White, M.P., Hattam, C., Pahl, S., King, H., Austen, M., 2019. Are some natural environments more psychologically beneficial than others? The importance of type and quality on connectedness to nature and psychological restoration. Environment and Behavior 51, 111–143.

Yang, K., Li, M., Liu, Y., Cheng, L., Huang, Q., Chen, Y., 2015. River detection in remotely sensed imagery using Gabor filtering and path opening. Remote Sensing 7, 8779–8802.

Yang, T., Jiang, S., Hong, Z., Zhang, Y., Han, Y., Zhou, R., Wang, J., Yang, S., Tong, X., Kuc, T.y., 2020. Sea-land segmentation using deep learning techniques for Landsat-8 OLI imagery. Marine Geodesy 43, 105–133.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., et al., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sensing of Environment 241, 111716.

Zhu, Q., Li, P., Li, Z., Pu, S., Wu, X., Bi, N., Wang, H., 2021. Spatiotemporal changes of coastline over the yellow river delta in the previous 40 years with optical and SAR remote sensing. Remote Sensing 13, 1940.

## List of Figures