



ORE Open Research Exeter

TITLE

Coastline detection in satellite imagery: A deep learning approach on new benchmark data

AUTHORS

Seale, C; Redfern, T; Chatfield, P; et al.

JOURNAL

Remote Sensing of Environment

DEPOSITED IN ORE

31 May 2022

This version available at

<http://hdl.handle.net/10871/129802>

COPYRIGHT AND REUSE

Open Research Exeter makes this work available in accordance with publisher policies.

A NOTE ON VERSIONS

The version presented here may differ from the published version. If citing, you are advised to consult the published version for pagination, volume/issue and date of publication

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

17 Catherine Seale^{a,b,*}, Thomas Redfern^a, Paul Chatfield^a, Chunbo Luo^b, Kari
18 Dempsey^a

^aUK Hydrographic Office, Taunton, TA1 2DN, Somerset, United Kingdom

^bInstitute for Data Science and Artificial Intelligence, University of Exeter, Exeter, EX4
4PY, Devon, United Kingdom

19 **Abstract**

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

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

*Corresponding author.

Preprint submitted to Remote Sensing of Environment
Email address: catherineseale@gmail.com (Catherine Seale)

April 3, 2022

³⁶ model training attention to the boundary between land and water. Through a
³⁷ hybrid quantitative and qualitative model assessment process we demonstrate
³⁸ that the model trained using our novel Sobel-edge loss function has greater
³⁹ sensitivity to fine-scale, narrow coastline features whilst possessing near top
⁴⁰ quantitative performance demonstrated by Categorical Cross-entropy.

⁴¹ The SWED dataset is published openly for use by the remote sensing and
⁴² machine learning communities, whilst the Sobel-edge loss is available for use
⁴³ in machine learning applications where sensitivity to boundary features is
⁴⁴ important.

⁴⁵ *Keywords:* Automated coastline extraction, Sentinel-2 satellite imagery,

⁴⁶ Deep Learning, Machine Learning, Labelled data, Loss Function

⁴⁷ *PACS:* 0000, 1111

⁴⁸ *2000 MSC:* 0000, 1111

⁴⁹ **1. Introduction**

⁵⁰ Coastal regions hold significant environmental, societal and economic
⁵¹ value (Wyles et al., 2019). Throughout history human populations have
⁵² been attracted to settle in coastal areas, due to the fertile soils, abundant
⁵³ food and opportunities for transport and trade in these regions (Edmonds
⁵⁴ et al., 2020). This is still the case today, with 41% of the global population
⁵⁵ living within 100 km of the coastline (Martinez et al., 2007). According to
⁵⁶ IPCC reports, coastal regions are particularly sensitive to the impacts of cli-
⁵⁷ mate change, and risks to these areas, from both natural and anthropogenic
⁵⁸ drivers, threaten both human populations and the ecosystems they rely upon
⁵⁹ (Wong et al., 2014). Detailed and up-to-date coastline morphology data, de-

60 fined here as the form or shape of the boundary between land and water,
61 underpins our understanding of coastline change and our ability to manage
62 its impacts over time (Burningham and French, 2017).

63 There are two main methodological sources of coastline morphology data,
64 in-situ measurements (Kuschnerus et al., 2021) and observations from re-
65 mote sensing technologies (Zhu et al., 2021). In-situ profiling or surveying
66 provides the most precise results but is only practicable for small regions.
67 The labour and costs involved render in-situ methods unfeasible for mapping
68 extensive areas or for repeated analyses. Indeed, in-situ measurement may
69 be impossible if a study area is remote, treacherous or inaccessible. Like-
70 wise, remotely sensed data or imagery captured by piloted flight or drone
71 are expensive to collect and impractical to apply at scale. In consequence,
72 the dominant approaches to mapping coastlines use satellite remote sensing
73 imagery and image processing techniques (Toure et al., 2019). The benefits
74 are numerous; satellites provide a rich time series of images as they revisit
75 an area every few days, they allow measurement without having to travel
76 to an area, and different instrumentation is available for a variety of use
77 cases. As such, satellite remote sensing is the only source of data by which
78 we might realistically continuously observe global coastline morphology. No-
79 table examples of free and open-access global satellite data are the European
80 Space Agency's Copernicus Programme (European Commission, 2013), and
81 the Landsat archive from NASA and the U.S. Geological Survey (Loveland
82 and Dwyer, 2012). These multi-petabyte datasets enable monitoring of the
83 Earth's surface over time (Li and Gong, 2016), presenting opportunities to
84 study the world's natural and anthropogenic environments at scale. There-

85 fore global, repeatable analysis of coastline morphology will be possible, if a
86 sufficiently automated, scalable and accurate method is devised for process-
87 ing coastal satellite images into coastline morphology data (Parente et al.,
88 2019).

89 This opportunity does however come with intrinsic limitations. Firstly,
90 the temporal record available for a satellite is fixed to the lifetime of its
91 operation, limiting potential for longer-term analysis. For example, data col-
92 lection by the European Space Agency's Sentinel-2 mission commenced in
93 2015, therefore analysis prior to this date is not possible using this sensor
94 alone. A longer archive is offered by the Landsat mission, with 30m resolu-
95 tion imagery available from 1999, however, this is of coarser spatial resolution
96 than Sentinel-2. Admitting the limitations of spatial resolution and tempo-
97 ral record with respect to global satellite data for the time being, the data
98 collected by these missions are proving suitable for study of coastlines. At
99 the time of writing, a small number of related analyses have been performed
100 on global satellite data. These examples include monitoring the extent of
101 sandy beaches (Luijendijk et al., 2018), tidal flats (Murray et al., 2018) and
102 mangrove (Bunting et al., 2018). But as yet, the lack of a viable, scalable
103 global method to observe and monitor coastline morphology still limits our
104 understanding of coastal zones. At present, the majority of studies perform-
105 ing coastline morphology detection restrict their focus to small local regions
106 (Uddin et al., 2020), with one notable exception from Bishop-Taylor et al.
107 (2019) that successfully extends detection to the Australian continent.

108 Local-scale, automated approaches to coastline morphology detection within
109 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

185 spatial quality.

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

207 **2. Data and Methods**

208 *2.1. Sentinel-2 Water Edges Dataset*

209 We introduce the Sentinel-2 Water Edges Dataset (SWED) for the devel-
210 opment and bench-marking of coastline detection methods. SWED contains
211 images captured by the European Space Agency’s Sentinel-2 satellites be-
212 tween 2017 and 2021. When selecting images, we constrained our search to
213 clear, cloud-free images by filtering the available catalog on the ‘cloudy pixel
214 percentage’ metadata, and visually inspecting the returned results. No addi-
215 tional pre-processing to the source imagery was applied. The selected images
216 are pre-allocated into train and test splits to support direct comparison of
217 models. We annotated the selected images to create dense, pixel-level labels
218 in two classes, ‘water’ and ‘non-water’. The distribution of images was man-
219 ually selected to ensure coverage of coastal environments that firstly span a
220 wide range of geographies, as shown in Figure 1, and secondly a variety of
221 coastline types at both high and low water conditions. During construction
222 of the dataset, we were unable to identify an authoritative source of global
223 coastline classification suitable for our needs, and therefore have created a
224 list of coastline types compiled from our research (see Table 1). In addi-
225 tion, we included examples of narrow fine-grained features such as jetties
226 and bridges. These features, whilst visible and recognisable to the human
227 eye, are often overlooked by algorithms by virtue of their small size. The
228 combination of images in SWED enables thorough testing of model sensi-
229 tivity to small features, capability at recognising various coastline types and
230 ability to geographically generalise. Here we explain the source data and
231 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.

232 *2.1.1. Source Data*

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

240 The highest spatial resolution of the Sentinel-2 Multi-Spectral Instrument
241 (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

²⁴² chine learning frameworks, any MSI bands at 20 or 60 metre resolution were
²⁴³ re-sampled to 10 metre resolution using two-dimensional nearest neighbour

²⁴⁴ interpolation. Consequently, one pixel within an image or array corresponds
²⁴⁵ to a geographical plan area of $10\text{m} \times 10\text{m}$ 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

²⁴⁶ *2.1.2. Training Set Annotation Process*

²⁴⁷ We created annotations using a semi-supervised clustering approach using
²⁴⁸ 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.

249 false colour image from spectral bands that display high contrast between
 250 water and non-water pixels. Sentinel-2 bands were selected for visualisation
 251 through trial and error to illicit the greatest visual contrast between water
 252 and non-water pixels. Whilst we did not find any band combination to be
 253 consistently successful across all training sites, the combinations of 8/11/4,
 254 8/4/3 and 4/3/1 were found to be a good starting point when rendered in
 255 red, green and blue channels respectively (see Table 2 for further details of
 256 these bands). Secondly, we applied k-means clustering to the rendered image.
 257 The number of clusters k was manually optimised to produce the best result
 258 for each individual image. Then, clusters were combined until two clusters
 259 remained, containing either water or non-water pixels. Lastly, we performed a
 260 visual comparison against high-resolution aerial imagery available in Google
 261 Earth and Bing Maps, and manually corrected any remaining mislabelled
 262 pixels to produce dense, pixel-level segmentation masks. An example image
 263 and mask are shown in Figure 2.

264 *2.1.3. Test Set Annotation Process*

265 A test set was created from a second batch of images that are geographi-
266 cally independent from those included in the training set. This is to create a
267 test set that will test the ability of any coastline extraction methodology to
268 generalise on imagery and geographies that are separate to that which was
269 used for training. This is a more robust test structure than testing on a sub-
270 set of the imagery that is used for training (López-Puigdollers et al., 2021)
271 as it tests method's ability to generalise to variation in coastline, geography
272 and Sentinel-2 scenes. The images in this set were curated to include a wide
273 range of coastal types, varied tidal states and fine-grained coastal features,
274 both natural and anthropogenic. Figure 1 shows the 49 selected geographic
275 locations sampled in the test set. Two Sentinel-2 images at each location, one
276 showing high water conditions, and one showing low water conditions, were
277 labelled. Creation of target segmentation masks for this set involved careful,
278 manual digitisation of the water/non-water boundary using the red, green,
279 blue and near-infrared bands on a 256 x 256 pixel subset of each image by an
280 experienced remote sensing analyst. The SWED test set therefore contains
281 98 image and segmentation mask pairs. The labelling effort applied to the
282 test set was more intensive than the training set, to increase confidence in
283 the overall position of the coastline within the test set labels.

284 *2.2. Convolutional Neural Network Architecture*

285 Convolutional Neural Networks consist of neuron layers that transform
286 an image input into a desired output (e.g., an image classification label or
287 segmented image mask). CNNs contain convolutional layers which apply
288 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 3×3 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 2×2 pool size. The decoder path has four corresponding blocks. Each decoder block contains an up-sampling layer of size 2×2 , followed by two convolutional layers, again using a 3×3 kernel, ELU activation, and batch

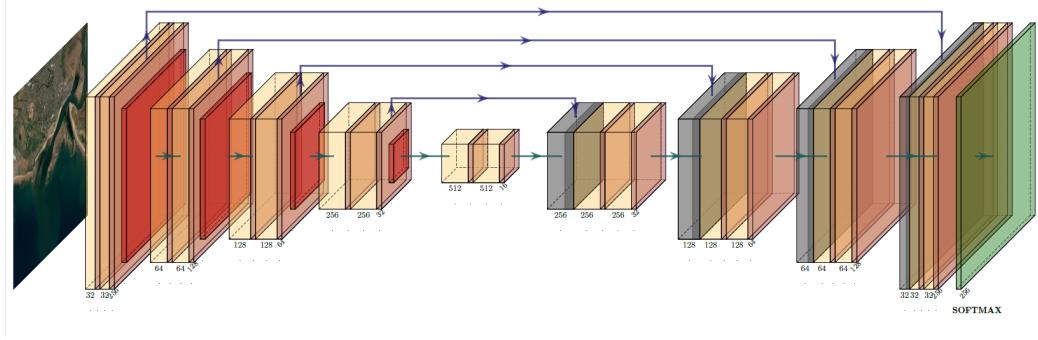


Figure 3: A schematic representation of the U-Net-based convolutional neural network architecture used to segment Sentinel-2 imagery into water and non-water classes. The same architecture was optimised multiple times with different loss functions, to compare the sensitivity of the resulting models to 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.

³¹⁴ normalisation layer. The last layer is a convolutional layer with 2 filters of
³¹⁵ size 1 x 1 with softmax activation.

³¹⁶ 2.3. Loss Functions

³¹⁷ 2.3.1. Standard Loss Functions

³¹⁸ A U-Net model optimised using cross-entropy loss provides an initial
³¹⁹ benchmark for segmentation performance on the SWED dataset. This model
³²⁰ represents the 'default' for coastline detection performance without any task-
³²¹ specific adaptations. Cross-entropy loss examines each pixel within an image
³²² individually, to compare the predicted class to the target class. Let p indicate
³²³ probability P that the pixel is of the positive class label Y for a pixel with a
³²⁴ 0 or 1:

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

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

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

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

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

$$SDC = \frac{2TP}{2TP + FN + FP} \quad (3)$$

339 Formulated as loss function, SDL is as follows:

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

340 The numerator of SDL is approximated by summing all the values from
 341 the element-wise multiplication between p and \hat{p} , and the denominator is

342 obtained by summing all the elements p and \hat{p} . A small value ϵ is added to
343 the denominator to avoid division by zero, and added to the numerator to
344 smooth the result (Planche and Andres, 2019). SDL, as with cross-entropy, is
345 calculated across all pixels, without any geographic or feature-specific focus.

346 *2.3.2. Edge-weighted Loss Functions*

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

355 Firstly, we trialled a novel loss function termed 'Sobel-edge loss'. This ap-
356 proach leverages the Sobel edge detection algorithm (Vincent and Folorunso,
357 2009) to extract the edges between segments. Performed on both predicted
358 and target segmentation, a loss is obtained by comparing the two sets of
359 edges. Our hope is that training will drive the predicted segments to have
360 increasingly similar edges to the target segments. We define Sobel-edge loss
361 as the Mean Square Error between edges detectable on targets and those
362 detectable on predictions, through the application of Sobel edge detection
363 filters. A Sobel edge detector filter uses two kernels that apply a convolution
364 operation to an input (Vincent and Folorunso, 2009). For an image A , the
365 two filters compute gradients along the x and y axes as below:

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

366 and

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

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

$$G = \sqrt{G_x^2 + G_y^2} \quad (7)$$

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

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

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

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

377 *2.3.3. Training Phase*

378 Four models were created, each using a different loss function but oth-
379 erwise identical in every aspect. Two of the models were optimised using
380 standard loss functions: Categorical cross-entropy loss and SDL. Two fur-
381 ther models were trained using Sobel-edge loss and W-SDL. The models
382 were trained on the SWED training set, using 23807 and 2661 256x256 pixel
383 patches for training and validation sets respectively. Training data was ran-
384 domly shuffled before input. Each model variant was trained for 50 epochs,
385 with early stopping if the validation loss did not improve after 10 epochs. The
386 learning rate was reduced by a factor of 0.1 when the validation loss plateau-
387 ed for more than 5 epochs. At the end of each epoch, the model weights were
388 saved if the validation loss improved, resulting in the final model weights
389 being those associated with the smallest validation loss over the course of
390 training.

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

394 *2.4. Model Assessment Methodology*

395 Model performance was evaluated in four ways:

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

402 dataset. Each test location was cross-referenced with the World Bank
 403 Land Boundaries dataset to define continent attributes (World Bank,
 404 2020).

- 405 3. Qualitative assessment of model sensitivity to fine-grained coastal fea-
 406 tures. Predictions on all 98 test images were visually assessed and
 407 patterns of consistent misclassification by the models were noted as
 408 well as the presence/absence of small-scale fine coastal detail at the
 409 segmentation boundary.
- 410 4. Positional accuracy of the defined water/non-water boundary.

411 *2.4.1. Quantitative evaluation metrics*

412 Quantitative evaluation metrics were calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

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

$$\text{Precision} = \frac{TP}{TP + FP} \quad (12)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (13)$$

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

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

$$\text{Jaccard} = \frac{TP}{TP + FP + FN} \quad (16)$$

413 where TP denotes true positives - correctly predicted water pixels. TN
 414 denotes true negatives - correctly predicted non-water pixels. FP denotes
 415 false positives - non-water pixels predicted to be water. FN denotes false
 416 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:

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

417 *2.4.2. Positional accuracy*

418 We determined the positional accuracy of predicted coastlines (the bound-
 419 ary between water and non-water pixels) using an adaptation of the “Trimap”
 420 method described by Kohli et al. (2008). The original Trimap method com-
 421 putes performance metrics within buffered regions around the target object
 422 boundary. We developed and used a variant to the original method that
 423 calculates the percentage of predicted boundary that falls within buffered
 424 regions of the target object boundary. Calculation of percentages within a
 425 tolerance allows us to describe results using statements of accuracy, such as
 426 “ $x\%$ of the predicted coastline is within y metres of the target boundary”.

427 To convert segmented images to a predicted coastline, we used an im-
 428 plementation of the marching squares algorithm (Lorensen and Cline, 1987)
 429 available in the scikit-image Python library (van der Walt et al., 2014). Val-
 430 ues between adjacent pixels were linearly interpolated and a line was drawn

431 along a contour of constant value. In this work, a segmented image consists
432 of two classes, water and non-water, represented by pixel values of either 0
433 or 1 and therefore the 0.5-valued contour was taken to be the coastline.

434 **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.

Loss FUNCTION	METRICS							
	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

435 *3.1. Quantitative metrics*

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

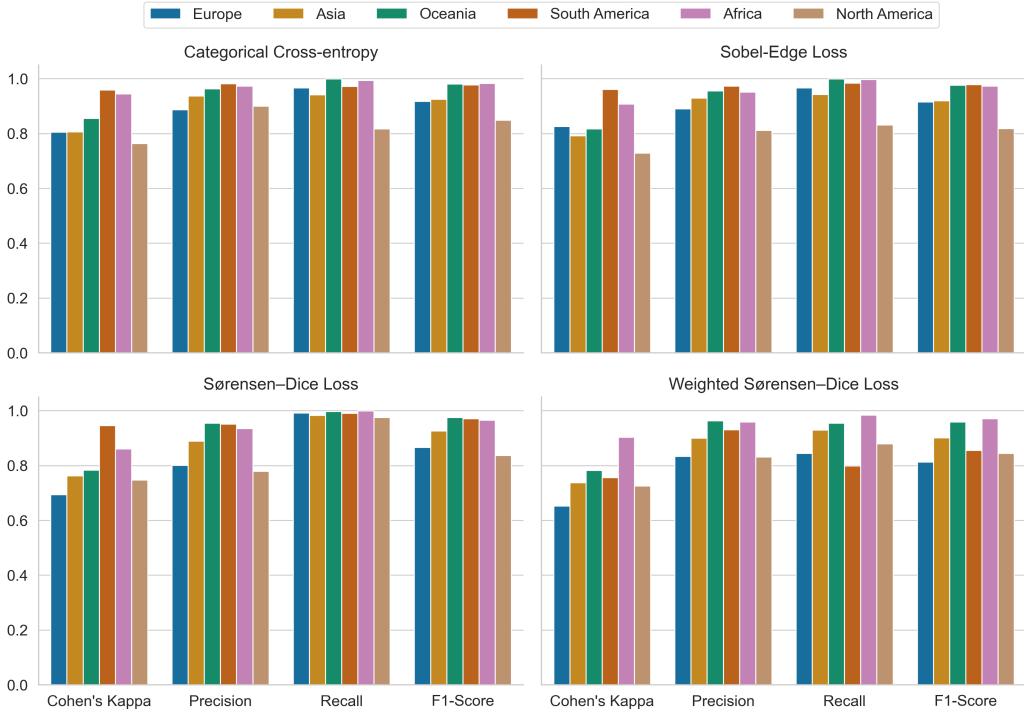


Figure 4: Performance metrics Cohen’s Kappa score, Precision, Recall and Mean F1-score for four models created using different loss functions, by geographic region.

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

443 *3.2. Generalisability to geographic region and tidal state*

444 Figure 4 examines the geographical variation of Cohen’s Kappa, preci-
 445 sion, recall and F1-score for each of the tested models. There is variety in
 446 model performance across the continents from which the SWED test set is
 447 sampled. For the F1-score statistic, a similar performance is recorded at con-
 448 tinental scale for the CXE and Sobel-edge models, with the model trained
 449 using Sørensen–Dice loss (“SDL Model”) exhibiting a similar behaviour apart

450 from the European performance, which is reduced. Unlike the other three
 451 models, the W-SDL model has a reduced performance in South America in
 452 comparison to the competing models. For Cohen’s Kappa score, most mod-
 453 els perform best in African and South American continents. The W-SDL
 454 model is the exception, once again the worst performer in South America.
 455 South America, Oceania, Asia and Africa have the highest precision and re-
 456 call across models, with Europe and North America tending to have a lower
 457 precision and recall across all models.

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

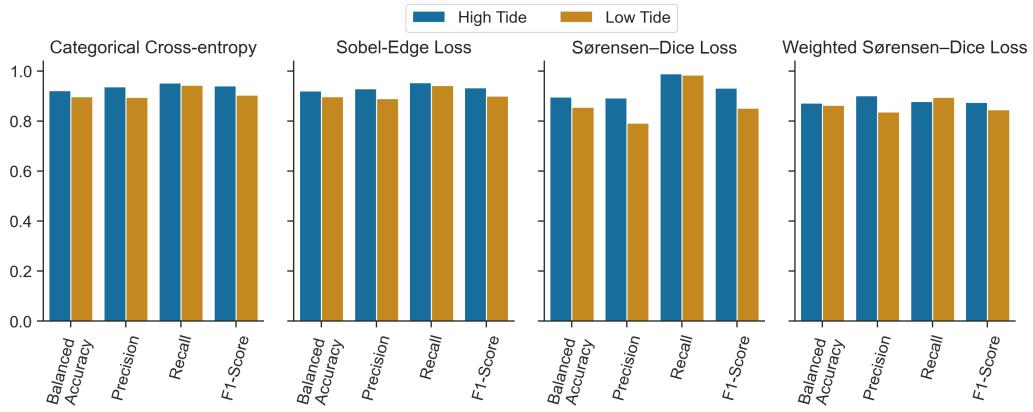


Figure 5: Quantitative metrics for four models created using different loss functions, when shown unseen images depicting conditions at either low or high tide.

462 *3.3. Sensitivity to small-scale geographic detail*

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

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

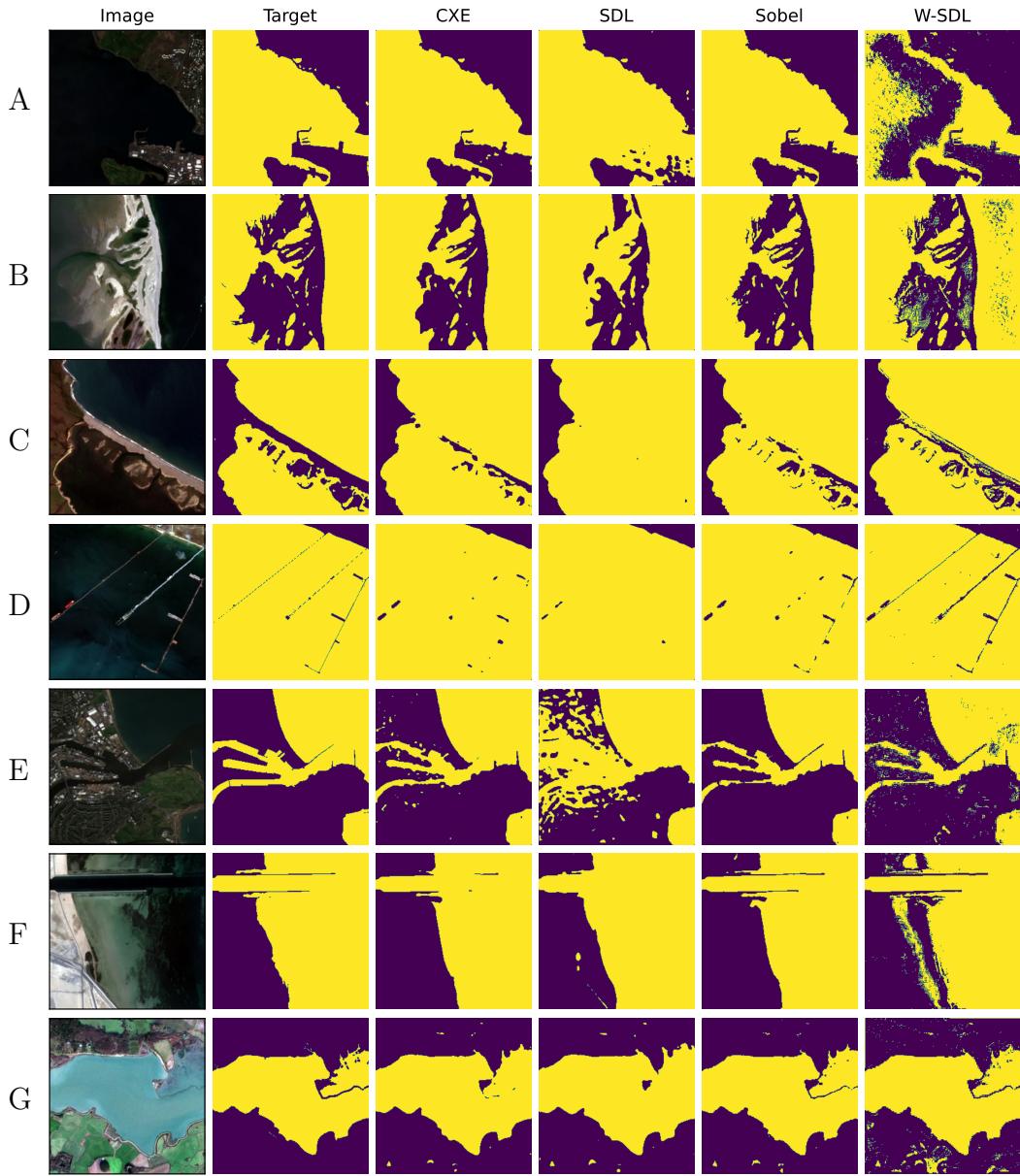


Figure 6: Visualised segmentation results. The first and second columns depict the source image and the target label respectively. The subsequent columns each depict the predicted segmentation map from a different model, identified by the loss function used during training, e.g., CXE: Categorical Cross-entropy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

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

493 *3.4. Positional accuracy of the defined coastline boundary*

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

505 **4. Discussion**

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

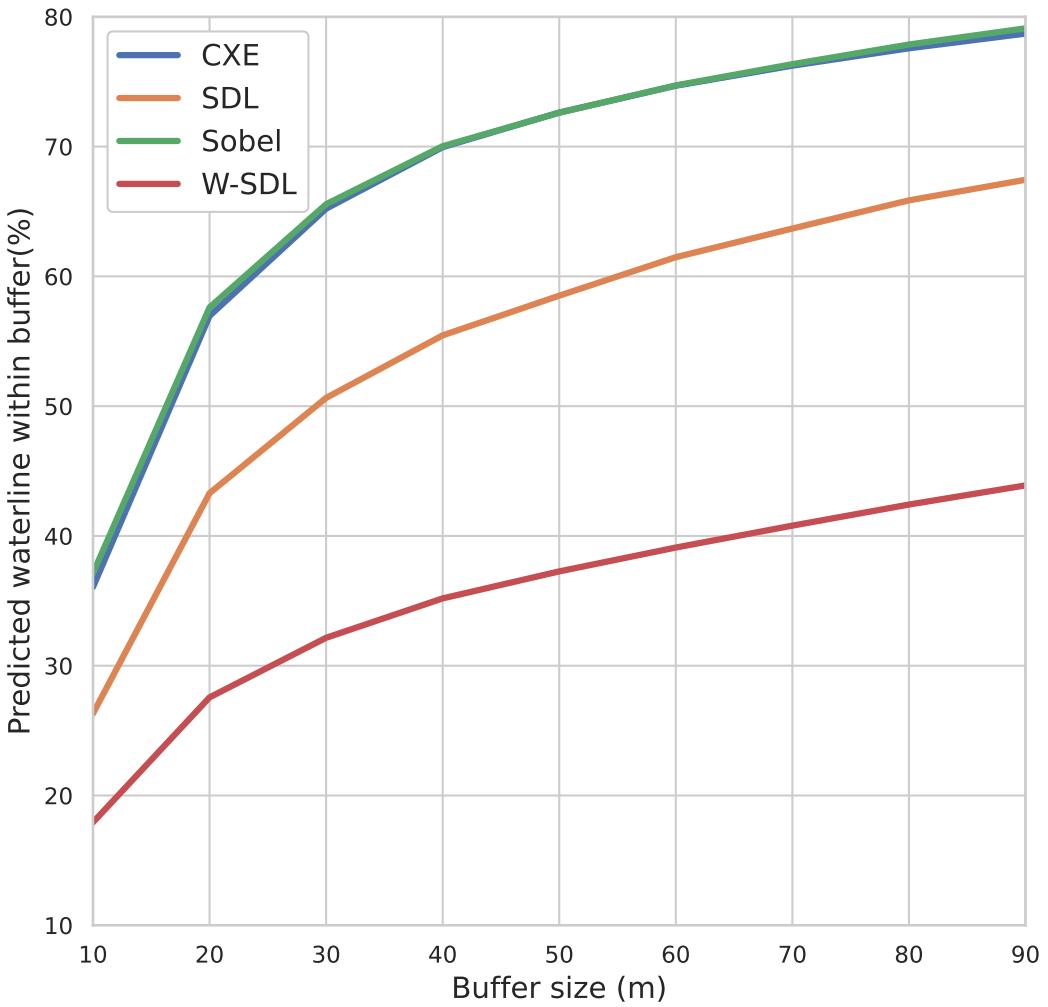


Figure 7: Comparison of what percentage of coastline generated by four different models under test is contained within a buffered region around the target coastal boundary. An increase in the y-axis is interpreted as producing coastlines with greater positional accuracy. The legend identifies the loss 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.

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

518 The qualitative analysis presented in Section 3.3 demonstrates that the
519 choice of loss function affects model performance and sensitivity of final
520 trained models to small coastal features. Models with similar quantitative
521 test performance are shown to possess different sensitivities to small scale
522 coastal detail, thus, future researchers should consider a full range of quanti-
523 tative and qualitative performance assessment methodologies when designing
524 and testing coastline extraction methodologies. The Sobel-edge loss proposed
525 by this study produced the best performing model by qualitative but not
526 quantitative analysis. Sobel-edge loss was particularly effective at the detec-
527 tion of narrow coastline features, such as bridges, breakwaters and jetties.
528 These small features are disproportionately important for their size, occur-
529 ring in ports and developing areas of human influence, but were often missed
530 by standard image segmentation loss functions. Sobel-edge loss was shown
531 to promote their persistence through the deep network. It was, however, dif-
532 ficult to quantify the ability of models to maintain fine detail. Such features
533 may only be a few pixels in size and consequently their presence or absence
534 did not greatly effect segmentation performance metrics or assessments of
535 coastline positional accuracy. To fully understand the nature of predictions,

536 visualisation and a hybrid qualitative and quantitative assessment was re-
537 quired. Relying on quantitative statistics alone would result in the choice of
538 a model that was insensitive to such fine coastal features.

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

545 Creating a single ML or DL model able to geographically generalise across
546 large regions is a challenge (Wang et al., 2017), that requires specific atten-
547 tion in the design of model testing strategies (Waldner and Diakogiannis,
548 2020; López-Puigdollers et al., 2021). In this study we compared model test
549 metrics at the continental scale to define a baseline of geographic variation in
550 model performance. Our analysis demonstrates that the performance of all
551 four tested models varies across the different continents and that there are
552 similarities between the CXE and Sobel-edge models. At this stage we are
553 unable to describe whether this variation in test performance is a result of
554 model sensitivity to coastal features, other landscape features within SWED
555 test imagery or some other phenomena that may lead to variation in test
556 imagery e.g. atmospheric interference as a result of aerosols or particulates.
557 Indeed, we have compared performance at a continental scale, but it could
558 be that a more fine-grained analysis at the country or landscape level may
559 provide further insight into variation in model performance. Whilst care was
560 taken to ensure that both training and test data contained examples of a di-

561 verse range of coastline types, we were unable to evaluate model performance
562 by coastline type in a quantitative manner. For this to be possible, pixel-level
563 labels of coastline type are required, but in this study only image-level labels
564 describing coastline types were prepared. An individual image very often
565 contained more than one type of coastline, e.g. an image may show a beach,
566 a rocky shore and a pier. Consequently, we were able to describe the total
567 count of examples depicting various coastline types, as detailed in Table 1,
568 but not which pixels within an image belong in which class. The creation of
569 pixel-level coastline class labels is therefore a recommendation for improving
570 SWED, as it would allow for detailed evaluation results with respect to coast-
571 line type. In lieu of an understanding of what is the optimal geographical
572 scale to understand variation in model performance, continental scale was
573 chosen as a pragmatic approach - as it was possible to cross reference our
574 test image locations with readily available data. Improving understanding
575 of the difficulties of generalising model performance geographically is rec-
576 ommended, so that future models may be more robustly tested and thus
577 deployed with confidence to areas outside of original model training. We also
578 compared model test metrics at different tidal states. The results indicated
579 that performance was greater on high-tide images than on low-tide images
580 of the same areas. An explanation may be the differing characteristics of a
581 low-tide line versus its high-tide counterpart. A low-tide line may have an
582 ambiguous boundary with no meaningful shape to it (Vos et al., 2019). In
583 a low-tide image, the landward side may be wet, for example saturated mud
584 or sand, providing reduced contrast between the two water/not water classes
585 and therefore be more difficult for classifiers to distinguish (Ryu et al., 2002;

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

592 A limitation of this study is the absence of in-situ measurements of coastal
593 position for creating training data and evaluating results which could offer
594 increased precision and confidence in labels. However, it is impractical to
595 take in-situ measurements for a study with global scope, and we were sub-
596 sequently constrained to visual interpretation of coastline position from im-
597 agery as a method of creating training and testing data. Variations in visual
598 interpretation may introduce uncertainty that could make it problematic to
599 assess performance. The problem of imperfect labels is unresolved and was
600 occasionally apparent in this study in cases where predictions appeared to be
601 more accurate than labels suggested. However, the visual interpretation of
602 remote sensing imagery is common in studies using image analysis and ma-
603 chine learning techniques with remote sensing imagery (Bunting et al., 2018;
604 Cheng et al., 2016), as it remains the only practical way of creating large,
605 geographically distributed training and testing sets. Another limitation may
606 be a lack of variation in training examples, which were constrained to clear,
607 cloud-free images and are therefore not representative of all Sentinel-2 im-
608 ages possible. As a result, segmentation models trained using these examples
609 will likely produce poor predictions if deployed indiscriminately on Sentinel-2
610 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

635 in test images. SWED and the Sobel-edge Loss may now be used to optimise
636 image segmentation networks for coastline detection.

637 **6. Data Access**

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

641 **7. Acknowledgements**

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

646 **References**

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

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

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

- 656 Bunting, P., Rosenqvist, A., Lucas, R.M., Rebelo, L.M., Hilarides, L.,
657 Thomas, N., Hardy, A., Itoh, T., Shimada, M., Finlayson, C.M., 2018.
658 The global mangrove watch — a new 2010 global baseline of mangrove
659 extent. *Remote Sensing* 10, 1669.
- 660 Burningham, H., French, J., 2017. Understanding coastal change using shore-
661 line trend analysis supported by cluster-based segmentation. *Geomorphol-*
662 *ogy* 282, 131–149.
- 663 Cao, W., Zhou, Y., Li, R., Li, X., 2020. Mapping changes in coastlines and
664 tidal flats in developing islands using the full time series of Landsat images.
665 *Remote Sensing of Environment* 239, 111665.
- 666 Cheng, D., Meng, G., Cheng, G., Pan, C., 2016. SeNet: Structured edge
667 network for sea–land segmentation. *IEEE Geoscience and Remote Sensing*
668 *Letters* 14, 247–251.
- 669 Chu, Z., Tian, T., Feng, R., Wang, L., 2019. Sea-land Segmentation with Res-
670 UNet and fully connected CRF, in: *Proceedings of the IEEE International*
671 *Geoscience and Remote Sensing Symposium*, pp. 3840–3843.
- 672 Clevert, D., Unterthiner, T., Hochreiter, S., 2016. Fast and accurate deep
673 network learning by Exponential Linear Units (ELUs), in: Bengio, Y., Le-
674 Cun, Y. (Eds.), *Proceedings of the 4th International Conference on Learn-*
675 *ing Representations*.
- 676 Edmonds, D., Caldwell, R., Brondizio, E., Siani, S., 2020. Coastal flooding
677 will disproportionately impact people on river deltas. *Nature Communi-*
678 *cations* 11, 4741.

- 679 European Commission, 2013. Legal notice on the use of Copernicus Sentinel
680 Data and Service Information. [https://sentinel.esa.int/documents/](https://sentinel.esa.int/documents/247904/690755/Sentinel_Data_Legal_Notice)
681 247904/690755/Sentinel_Data_Legal_Notice. Accessed 2021-03-17.
- 682 Galeone, P., 2019. Hands-On Neural Networks with TensorFlow 2.0. Packt
683 Publishing.
- 684 Heene, G., Gautama, S., 2000. Optimisation of a coastline extraction al-
685 gorithm for object-oriented matching of multisensor satellite imagery, in:
686 IEEE International Geoscience and Remote Sensing Symposium, pp. 2632–
687 2634.
- 688 Karantzalos, K.G., Argialas, D., Georgopoulos, A., 2002. Towards auto-
689 matic detection of coastlines from satellite imagery, in: 14th International
690 Conference on Digital Signal Processing Proceedings, pp. 897–900.
- 691 Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on con-
692 volutional neural networks (CNN) in vegetation remote sensing. ISPRS
693 Journal of Photogrammetry and Remote Sensing 173, 24–49.
- 694 Klinger, T., Ziems, M., Heipke, C., Hannover Schenke, H.W., Ott, H., 2012.
695 Automated extraction of the Antarctic coastline using snakes. Interna-
696 tional Archives of the Photogrammetry, Remote Sensing and Spatial In-
697 formation Sciences 38.
- 698 Kohli, P., Ladicky, L., Torr, P.H.S., 2008. Robust higher order potentials for
699 enforcing label consistency, in: Preceedings of the 2008 IEEE Conference
700 on Computer Vision and Pattern Recognition, pp. 1–8.

- 701 Kuschnerus, M., Lindenbergh, R., Vos, S., 2021. Coastal change patterns
702 from time series clustering of permanent laser scan data. *Earth Surface*
703 *Dynamics* 9, 89–103.
- 704 LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–
705 444.
- 706 Li, R., Liu, W., Yang, L., Sun, S., Hu, W., Zhang, F., Li, W., 2018. Deep-
707 UNet: A deep fully convolutional network for pixel-level sea-land segmen-
708 tation. *IEEE Journal of Selected Topics in Applied Earth Observations*
709 and *Remote Sensing* 11, 3954–3962.
- 710 Li, W., Gong, P., 2016. Continuous monitoring of coastline dynamics in
711 western Florida with a 30-year time series of Landsat imagery. *Remote*
712 *Sensing of Environment* 179, 196 – 209.
- 713 Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for
714 dense object detection, in: *Proceedings of the 2017 IEEE International*
715 *Conference on Computer Vision*, pp. 2999–3007.
- 716 Liu, H., Jezek, K.C., 2004. Automated extraction of coastline from satellite
717 imagery by integrating canny edge detection and locally adaptive thresh-
718 olding methods. *International Journal of Remote Sensing* 25, 937–958.
- 719 López-Puigdollers, D., Mateo-García, G., Gómez-Chova, L., 2021. Bench-
720 marking deep learning models for cloud detection in Landsat-8 and
721 Sentinel-2 Images. *Remote Sensing* 13, 992.
- 722 Lorensen, W.E., Cline, H.E., 1987. Marching cubes: A high resolution 3d

- 723 surface construction algorithm, in: Proceedings of the 14th Annual Con-
724 ference on Computer Graphics and Interactive Techniques, Association for
725 Computing Machinery. pp. 163—169.
- 726 Loveland, T.R., Dwyer, J.L., 2012. Landsat: Building a strong future. Re-
727 mote Sensing of Environment 122, 22–29.
- 728 Luijendijk, A., Hagenaars, G., Ranasinghe, R., Baart, F., Donchyts, G.,
729 Aarninkhof, S., 2018. The state of the world’s beaches. Scientific Reports
730 8, 6641.
- 731 Martinez, M., Intralawan, A., Vázquez, G., Pérez-Maqueo, O., Sutton, P.,
732 2007. The coasts of our world: Ecological, economic and social importance.
733 Ecological Economics 63, 254–272.
- 734 Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-Net: Fully convolutional neu-
735 ral networks for volumetric medical image segmentation, in: Proceedings
736 of the 2016 Fourth International Conference on 3D Vision, pp. 565–571.
- 737 Murray, N., Phinn, S., DeWitt, M., Ferrari, R., Johnston, R., Lyons, M.,
738 Clinton, N., Thau, D., Fuller, R.A., 2018. The global distribution and
739 trajectory of tidal flats. Nature 565, 222–225.
- 740 Oehmcke, S., Thrysøe, C., Borgstad, A., Salles, M.A.V., Brandt, M., Gieseke,
741 F., 2019. Detecting hardly visible roads in low-resolution satellite time
742 series data, in: IEEE International Conference on Big Data (Big Data),
743 pp. 2403–2412.

- 744 Paravolidakis, V., Ragia, L., Moirogiorgou, K., Zervakis, M., 2018. Automatic coastline extraction using edge detection and optimization procedures. *Geosciences* 8, 407.
- 747 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.
- 750 Planche, B., Andres, E., 2019. *Hands-On Computer Vision with TensorFlow* 2. Packt Publishing Ltd.
- 752 QGIS Development Team, 2021. QGIS Geographic Information System.
753 QGIS Association. Accessed 2020-08-11.
- 754 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.
756 pp. 234–241.
- 758 Ryu, J.H., Won, J.S., Min, K.D., 2002. Waterline extraction from Landsat
759 TM data in a tidal flat: A case study in Gomso Bay, Korea. *Remote
760 Sensing of Environment* 83, 442–456.
- 761 Shamsolmoali, P., Zareapoor, M., Wang, R., Zhou, H., Yang, J., 2019. A
762 novel deep structure U-Net for sea-land segmentation in remote sensing
763 images. *IEEE Journal of Selected Topics in Applied Earth Observations
764 and Remote Sensing* 12, 3219–3232.
- 765 Sudre, C., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J., 2017. Generalised Dice overlap as a deep learning loss function for highly unbalanced

- 767 segmentations, in: Deep learning in medical image analysis and multi-
768 modal learning for clinical decision support, pp. 240–248.
- 769 Sultana, F., Sufian, A., Dutta, P., 2020. Evolution of image segmentation
770 using deep convolutional neural network: A survey. Knowledge-Based
771 Systems 201-202, 106062.
- 772 Toure, S., Diop, O., Kpalma, K., Maiga, A.S., 2019. Shoreline detection
773 using optical remote sensing: A review. ISPRS International Journal of
774 Geo-Information 8.
- 775 Uddin, K., Khanal, N., Chaudhary, S., Maharjan, S., Thapa, R.B., 2020.
776 Coastal morphological changes: Assessing long-term ecological transfor-
777 mations across the northern Bay of Bengal. Environmental Challenges 1,
778 100001.
- 779 Vincent, O.R., Folorunso, O., 2009. A descriptive algorithm for Sobel im-
780 age edge detection, in: Proceedings of the 2009 Informing Science IT
781 Education Conference.
- 782 Vos, K., Harley, M.D., Splinter, K.D., A, S.J., Turner, I.L., 2019. Sub-annual
783 to multi-decadal shoreline variability from publicly available satellite im-
784 agery. Coastal Engineering 150, 160 – 174.
- 785 Waldner, F., Diakogiannis, F.I., 2020. Deep learning on edge: Extracting
786 field boundaries from satellite images with a convolutional neural network.
787 Remote Sensing of Environment 245, 111741.
- 788 van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner,

- 789 J.D., Yager, N., Gouillart, E., Yu, T., and the Scikit-image contributors,
790 2014. Scikit-image: image processing in Python. PeerJ 2, 453.
- 791 Wang, R., Camilo, J., Collins, L.M., Bradbury, K., Malof, J.M., 2017. The
792 poor generalization of deep convolutional networks to aerial imagery from
793 new geographic locations: an empirical study with solar array detection,
794 in: Proceedings of the 2017 IEEE Applied Imagery Pattern Recognition
795 Workshop, pp. 1–8.
- 796 Wong, P., Losada, I., Gattuso, J., Hinkel, J., Khattabi, A., McInnes, K.,
797 Saito, Y., Sallenger, A., 2014. Coastal systems and low-lying areas, in:
798 Field, C., Barros, V., Dokken, D., Mach, K., Mastrandrea, M., Bilir, T.,
799 Chatterjee, M., Ebi, K., Estrada, Y., Genova, R., Girma, B., Kissel, E.,
800 Levy, A., MacCracken, S., Mastrandrea, P., White, L. (Eds.), Climate
801 Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and
802 Sectoral Aspects. Contribution of Working Group II to the Fifth Assess-
803 ment Report of the Intergovernmental Panel on Climate Change. Cam-
804 bridge University Press, Cambridge, United Kingdom and New York, NY,
805 USA, pp. 361–409.
- 806 World Bank, 2020. World Bank Official Boundaries. <https://datacatalog.worldbank.org/search/dataset/0038272/>
807 **World-Bank-Official-Boundaries**. Accessed 17-09-2020.
- 808
- 809 Wyles, K.J., White, M.P., Hattam, C., Pahl, S., King, H., Austen, M., 2019.
810 Are some natural environments more psychologically beneficial than oth-
811 ers? The importance of type and quality on connectedness to nature and
812 psychological restoration. Environment and Behavior 51, 111–143.

- 813 Yang, K., Li, M., Liu, Y., Cheng, L., Huang, Q., Chen, Y., 2015. River detec-
814 tion in remotely sensed imagery using Gabor filtering and path opening.
815 Remote Sensing 7, 8779–8802.
- 816 Yang, T., Jiang, S., Hong, Z., Zhang, Y., Han, Y., Zhou, R., Wang, J., Yang,
817 S., Tong, X., Kuc, T.y., 2020. Sea-land segmentation using deep learning
818 techniques for Landsat-8 OLI imagery. Marine Geodesy 43, 105–133.
- 819 Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang,
820 Q., Wang, J., et al., 2020. Deep learning in environmental remote sens-
821 ing: Achievements and challenges. Remote Sensing of Environment 241,
822 111716.
- 823 Zhu, Q., Li, P., Li, Z., Pu, S., Wu, X., Bi, N., Wang, H., 2021. Spatiotemporal
824 changes of coastline over the yellow river delta in the previous 40 years with
825 optical and SAR remote sensing. Remote Sensing 13, 1940.

826 **List of Figures**

- 827 1 The distribution of labelled Sentinel-2 satellite imagery con-
828 tained in the Sentinel-2 Water Edges Dataset (SWED). Anno-
829 tations in two classes, water and non-water, were created for
830 Sentinel-2 scenes at the highlighted coastal areas. 16 training
831 and 49 testing locations are shown in green and red, respectively. 10
- 832 2 An example Sentinel-2 scene and corresponding segmentation
833 mask taken from the Sentinel-2 Water Edges Dataset (SWED),
834 showing pixels classified into 'non-water' and 'water' classes. . 13

835	3	A schematic representation of the U-Net-based convolutional neural network architecture used to segment Sentinel-2 im- agery 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 fea- tures. Convolutional, batch normalisation, max-pooling, up- sampling and softmax layers are shown in yellow, brown, red, grey and green respectively. Arrows denote skip connections. . .	16
843	4	Performance metrics Cohen's Kappa score, Precision, Recall and Mean F1-score for four models created using different loss functions, by geographic region.	24
846	5	Quantitative metrics for four models created using different loss functions, when shown unseen images depicting conditions at either low or high tide.	25
849	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.	27

