

Article

# SNOWED: Automatically Constructed Dataset of Satellite Imagery for Water Edge Measurements

Gregorio Andria, Marco Scarpetta \*, Maurizio Spadavecchia \*, Paolo Affuso and Nicola Giaquinto 

Department of Electrical and Information Engineering, Polytechnic University of Bari, Via E. Orabona 4, 70125 Bari, Italy; nicola.giaquinto@poliba.it (N.G.)

\* Correspondence: marco.scarpetta@poliba.it (M.S.); maurizio.spadavecchia@poliba.it (M.S.)

**Abstract:** Monitoring the shoreline over time is essential to quickly identify and mitigate environmental issues such as coastal erosion. Monitoring using satellite images has two great advantages, i.e., global coverage and frequent measurement updates; but adequate methods are needed to extract shoreline information from such images. To this purpose, there are valuable non-supervised methods, but more recent research has concentrated on deep learning because of its greater potential in terms of generality, flexibility, and measurement accuracy, which, in contrast, derive from the information contained in large datasets of labeled samples. The first problem to solve, therefore, lies in obtaining large datasets suitable for this specific measurement problem, and this is a difficult task, typically requiring human analysis of a large number of images. In this article, we propose a technique to automatically create a dataset of labeled satellite images suitable for training machine learning models for shoreline detection. The method is based on the integration of data from satellite photos and data from certified, publicly accessible shoreline data. It involves several automatic processing steps, aimed at building the best possible dataset, with images including both sea and land regions, and correct labeling also in the presence of complicated water edges (which can be open or closed curves). The use of independently certified measurements for labeling the satellite images avoids the great work required to manually annotate them by visual inspection, as is done in other works in the literature. This is especially true when convoluted shorelines are considered. In addition, possible errors due to the subjective interpretation of satellite images are also eliminated. The method is developed and used specifically to build a new dataset of Sentinel-2 images, denoted SNOWED; but is applicable to different satellite images with trivial modifications. The accuracy of labels in SNOWED is directly determined by the uncertainty of the shoreline data used, which leads to sub-pixel errors in most cases. Furthermore, the quality of the SNOWED dataset is assessed through the visual comparison of a random sample of images and their corresponding labels, and its functionality is shown by training a neural model for sea–land segmentation.

**Keywords:** satellite monitoring; deep learning; sea–land segmentation; shoreline detection; AI-based measurements; automatic labeled dataset construction; Sentinel-2; benchmark datasets



**Citation:** Andria, G.; Scarpetta, M.; Spadavecchia, M.; Affuso, P.; Giaquinto, N. SNOWED: Automatically Constructed Dataset of Satellite Imagery for Water Edge Measurements. *Sensors* **2023**, *23*, 4491. <https://doi.org/10.3390/s23094491>

Academic Editor: Carmine Serio

Received: 22 February 2023

Revised: 28 April 2023

Accepted: 3 May 2023

Published: 5 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Coastlines are crucial ecosystems with both environmental and economic significance, as nearly half of the world's population lives within 100 km of the sea [1]. These areas face various threats, including, fishing, pollution, shipping, and various consequences of climate change [2–4], making it imperative to monitor them for early detection of potential issues such as coastal erosion, that can cause harm to the environment and human settlements. Coastal monitoring can include detecting microplastics [5,6], and monitoring seagrasses [7,8], water quality [9,10], and antibiotics pollution [11,12] among others. Monitoring using in situ measurements is the most precise, but it can be costly and time-consuming, especially for large areas and/or frequent measurements. Remote sensing is an alternative solution that has evolved from aerial imagery taken from aircraft

for the use of Unmanned Aerial Vehicles (UAVs) and Unmanned Underwater Vehicles (UUVs). Such methods of remote sensing offer advantages over in situ measurements, but still require extensive human intervention and specialized technologies.

More recently, satellite imagery has become a promising additional monitoring technique. Satellite data are characterized by global coverage and high temporal resolution and are often publicly accessible. Sentinel-2 and Landsat 8 are two of the most used Earth observation satellites, capturing multispectral images of the Earth's surface with a resolution of up to 10 m. They provide, for a wide range of users including governments, academic institutions, and private companies, valuable information for monitoring changes in land cover and land use, as well as for detecting and mapping natural hazards [13–15]. The revisiting time of a few days enables near real-time monitoring of dynamic events on the earth's surface. The increasing demand for high-quality earth observation data makes Sentinel-2 and Landsat 8 expected to remain key players in the earth observation satellite market in the future.

Lines delimiting water regions may be extracted from satellite images using traditional signal processing methods. Even in the AI era, these methods are valuable and often optimal tools to extract information from signals and images [16–21]. As regards the specific topic of coastline monitoring, edge detection algorithms are used in [16] for Sentinel Synthetic Aperture Radar (SAR) images, obtaining an extracted coastline with a mean distance of 1 pixel from the reference shoreline, measured through in situ analysis. In [22] coastline is achieved from very-high-resolution Pléiades imagery using the Normalized Difference Water Index (NDWI), which is one of the most popular techniques for automatic coastline extraction. NDWI is also used in [23], but, in this case, results are improved by using repeated measurements and adaptive thresholding. Another example of traditional signal processing for coastline detection is [18], where shorelines were extracted from multispectral images using a new water-land index that enhances the contrast between water and land pixels. Yet another example is [19], where unsupervised pixel classification is used for extracting shorelines from high-resolution satellite images.

Sentinel-2 satellite images are frequently employed for the purpose of coastline extraction, owing to their high spatial resolution and multispectral capabilities. In [24], shoreline changes in the Al Batinah region of Oman and the impact of Cyclone Kyarr are analyzed using Sentinel-2 images and the Digital Shoreline Analysis System (DSAS). In [25], the effectiveness of MODIS, Landsat 8, and Sentinel-2 in measuring regional shoreline changes is compared. Shorelines are extracted, again, with the DSAS and Sentinel-2 is identified as the most effective source of satellite images due to its higher spatial resolution. Another tool proposed for shoreline extraction is the SHOREX system [26,27]. It is able to automatically define the instantaneous shoreline position at a sub-pixel level from Landsat 8 and Sentinel 2 images. In [28], shoreline changes associated with volcanic activity in Anak Krakatau, Indonesia, are analyzed using a NDWI-based method on Sentinel-2 multispectral imagery.

In more recent years, semantic interpretation of images is being performed more and more by means of supervised machine learning, i.e., deep neural networks (DNN), due to its successful applicability in very different fields and to very different kinds of images, and shoreline extraction from satellite imagery is no exception [29,30]. The well-known U-Net architecture [31], in particular, is often used for effective DNN-based coast monitoring on a global scale [32–40]. Different types of satellite images have been used for this purpose, including Sentinel-1 SAR images [32–34], Landsat 8 and Gaofen-1 multispectral images [35,40], and true color images (TCI) from Google Earth [36–39]. In [41], eight deep learning models, including different variations of U-Net, are used for coastline detection, and their performances are compared. Other kinds of deep neural networks have also been proposed for coastline detection. In [42], ALOS-2 SAR images are analyzed using a densely connected neural network with two hidden layers. A multi-task network which includes both a sea–land segmentation and a sea–land boundary detection module, named BS-Net is instead proposed in [43].

The key requirement to successful supervised machine learning is, of course, the availability of large datasets of accurately labeled samples. The problem with coastline detection from satellite images is that datasets of appropriate size are not common. A possible solution is to build synthetic datasets, i.e., a collection of artificially generated realistic images, produced by a computer program together with the associated “exact” labels. Synthetic datasets have been built and used successfully in many applications [44–46], but their construction is unpractical for satellite images, which contain complex patterns difficult to reproduce realistically with computer graphics. This is true for TCI images and even more for images in other spectral bands. Manual labeling, based on visual interpretation of TCIs, is a long and boring task, but usable effectively as long as the shorelines are comparatively simple, e.g., with sea and land separated by a single line; when many images in the dataset have elaborate shorelines (as in the example that will be shown later), it becomes impractical and very burdensome.

The present paper, extending the preliminary research in [47] (where a much smaller dataset is obtained), presents a method for automatically building a dataset of labeled satellite images for sea–land segmentation and shoreline detection. The method is based on the usage of publicly available shoreline data, together with publicly available satellite images. In particular, the method is developed to use shoreline data from the National Oceanic and Atmospheric Administration (NOAA) and satellite images from Copernicus Sentinel 2 project, obtaining the “Sentinel2-NOAA Water Edges Dataset” (SNOWED) [48], whose main features are the following.

- SNOWED is constructed with a fully automatic algorithm, without human intervention or interpretations.
- SNOWED is annotated using certified shoreline measurements.
- SNOWED contains satellite images of different types of coasts, located in a wide geographical area, including images related to very elaborate shorelines.

One intrinsic drawback of the automatic generation process is that some satellite images can contain water regions not included in the shoreline data used for labeling it and, hence, may have an incomplete label. This problem is however identified and handled as described in Section 4.1.

With respect to other datasets of this type that have been proposed in recent years, the method presented in the paper to generate the SNOWED dataset is characterized by some innovative aspects. Datasets found in the literature are all based on the visual interpretation of satellite images and require therefore a strong human effort for labeling. Furthermore, the accuracy of sea/land segmentation labels depends directly on the quality of satellite images selected for the dataset and on how well they can be visually interpreted by humans. The methodology designed and implemented for this work uses instead independent measurements to automatically generate the labeled dataset, without any human intervention. This translates both to avoidance of tiresome human work and to the generation of sea/land labels having known and very low uncertainty. In addition, by using the proposed method, the source of satellite images (which is the Sentinel 2 project in the case of SNOWED) can be easily changed while using still the same shoreline measurements, leading to broader possibilities of application.

The paper is organized as follows. In Section 2, a review of the available datasets of satellite images for sea/land segmentation tasks is presented. In Section 3, the automatic dataset generation procedure is presented. In Section 4, results of the application of the proposed generation method are reported, together with quality assessment results. In Section 5 are the conclusions.

## 2. Publicly Available Datasets of Satellite Images for Sea/Land Segmentation

The aim of this section is to illustrate the already available public datasets developed for training deep learning models for sea/land segmentation. Particular attention is dedicated to the general characteristics of the provided datasets and to the generation process used to obtain them. This is useful to understand the novelty and relevance of the

proposed work, and to conveniently compare the proposed SNOWED dataset with the other available alternatives.

Two major features are considered for each dataset: the number of samples containing both water and land pixels, and the source of labeling information. These features are indeed the only ones strictly related to the effectiveness of a dataset for training a neural network. We consider the number of images containing both land and sea, rather than the absolute number of images, because samples containing only one class can be trivially extracted from large areas that are known to contain only sea or only land. Besides, we highlight that the source of labeling information determines the accuracy of the labels, and hence the accuracy of models trained using the dataset.

### 2.1. Water Segmentation Data Set (QueryPlanet Project)

The water segmentation data set [49] has been created as a part of the QueryPlanet project, which has been funded by the European Space Agency (ESA). The dataset is composed of satellite images of size  $64 \times 64$  from the Sentinel-2 Level-1C product. Each of them has been manually labeled by volunteer users of a collaborative web app. Volunteers were prompted with an initial label obtained by calculating the NDWI [50] and had to visually compare it with the corresponding satellite TCI and correct eventual discrepancies based on their interpretation of the image. The online labeling campaign led to the creation of 7671 samples, but only 5177 of them contain both sea and land pixels.

### 2.2. Sea–Land Segmentation Benchmark Dataset

The dataset proposed in [51] contains labeled Landsat-8 Operational Land Imager (OLI) satellite images, of different types of Chinese shorelines: sandy, muddy, artificial and rocky coasts. The labels of the dataset are obtained through a multi-step human annotation procedure. First, Landsat-8 OLI images with less than 5% cloud cover are selected along the Chinese shoreline. These images are pre-processed by applying radiometric calibration and atmospheric correction and then are manually annotated by dividing all their pixels into two classes: sea and land. Finally, satellite images are cut into small patches and each patch is checked to remove the defective ones (e.g., blank and cloud-covered patches). At the end of the procedure, 3361 images of size  $512 \times 512$  are obtained, but only 831 of them contain both classes.

### 2.3. YTU-WaterNet

The YTU-WaterNet dataset proposed in [52] contains Landsat-8 OLI images too. The dataset is created starting from 63 Landsat-8 OLI full-frames containing coastal regions of Europe, South and North America, and Africa. Only the blue, red and near-infrared bands are used for the samples of the dataset to reduce the dataset size and the computational load needed for training operations. The satellite images are cut into  $512 \times 512$  patches and binary segmented by exploiting OpenStreetMap (OSM) water polygons data [43]. OSM data is created by volunteers based on their geographical knowledge of the area or on visual interpretation of satellite images. This data is available as vector polygons, which are then converted to raster images representing the water regions of the sample. Finally, a filtering operation is performed to eliminate cloud-covered samples and samples with only one class, while samples with a mismatching label are identified and eliminated by visual inspection. The YTU-WaterNet dataset contains 1008 images.

### 2.4. Sentinel-2 Water Edges Dataset (SWED)

The most recent dataset is the Sentinel-2 Water Edges Dataset (SWED), proposed in a research work supported by the UK Hydrographic Office [53]. SWED uses Sentinel-2 Level-2A imagery, semantically annotated through a semi-automatic procedure. The first step of the dataset creation process is the selection of Sentinel-2 images between 2017 and 2021. Only clear and cloud-free images are selected, by filtering on the 'cloudy pixel percentage' metadata associated with each image, and then by visually inspecting the obtained search

results. Furthermore, images are manually selected to cover a wide variety of geographical areas and types of coasts. A water/non-water segmentation mask is therefore created for each of the selected Sentinel-2 images. First, a false color image with visually good contrast between water and non-water pixels is searched by trial and error among those that can be obtained by rendering different combinations of Sentinel-2 bands in the RGB channels. The selected combination of bands is not the same for each image, although three combinations are found to be a good starting point. Secondly, a manually refined k-means-based procedure is applied to the rendered false color images to collect their pixels into two clusters, corresponding to water and non-water regions. Finally, segmentation masks are manually corrected by visual comparison against high-resolution aerial imagery available on Google Earth and Bing Maps. This imagery is, however, obtained as a composition of multiple images acquired on different and not precisely known dates, and therefore in some cases can represent inaccurately the actual state of the coasts at the Sentinel-2 acquisition time. The SWED dataset contains 26,468 images of size  $256 \times 256$ , cut from the annotated Sentinel-2 full-tiles, but only 9013 of them contain both classes.

### 2.5. Summary of the Characteristics of the Already Available Datasets, and of the New SNOWED Dataset

The described datasets are the result of a very intense effort and provide solid solutions for training and benchmarking machine learning models for shoreline recognition. Of course, a larger number of samples, or another dataset that can be used together with them, is desirable. Another improvable feature is the labeling process accuracy: a specific quality assurance on the shoreline labels would be a clear plus.

The methodology described in this work aims precisely at these goals: providing further samples useful for training neural models for satellite coastline measurements, along with labels coming from certified coastline measurements.

Table 1 summarizes the main characteristics of the four datasets in the literature described in this section, and those of the SNOWED dataset obtained with the procedure illustrated in the present work. The number of images reported in Table 1 refers to images containing both sea and land classes. As can be seen, the image size is not the same for all the datasets and therefore a conversion is needed to directly compare the number of images of the two datasets. For example, the 1008  $512 \times 512$  images of YTU-WaterNet correspond to 4032  $256 \times 256$  images.

**Table 1.** Main characteristics of the four examined datasets. The same characteristics for the new SNOWED dataset are included for comparison. They highlight that SNOWED (i) is compatible with, and adds up, to SWED; (ii) uses NOAA measurements, instead of human interpretation of images.

| Dataset ID                                   | N. of Images | Image Size       | Source of Coastline Data   |
|--|--------------|------------------|--|
| QueryPlanet [49]                             | 5177         | $64 \times 64$   | Human interpretation of TCI images   |
| Sea–land segmentation benchmark dataset [51] | 831          | $512 \times 512$ | Human interpretation of TCI images   |
| YTU-WaterNet [52]                            | 1008         | $512 \times 512$ | Human-generated OpenStreetMap water polygons data  |
| SWED [53]                                    | 9013         | $256 \times 256$ | Human interpretation of high-resolution aerial imagery available in Google Earth and Bing Maps |
| SNOWED                                       | 4334         | $256 \times 256$ | U.S. NOAA shoreline measurements   |

### 3. Data and Methods

The methodology presented in this paper consists in combining publicly available satellite images and shoreline data. This is a non-trivial task since many preprocessing operations and quality checks are needed to obtain accurately annotated samples. The methodology, and the involved processing, are illustrated with the concrete construction of a dataset, where the source of satellite imagery is the Level-1C data product of the

Sentinel-2 mission, and the source of shoreline data is the Continually Updated Shoreline Product (CUSP).

### 3.1. Data Sources

#### 3.1.1. Sentinel-2 Satellite Imagery

The Sentinel-2 mission consists of a constellation of two satellites for Earth observation, phased at 180° to each other to provide a revisit period of at most 5 days. The satellites are equipped with the MultiSpectral Instrument (MSI) that acquires images in 13 spectral bands, with spatial resolutions up to 10 m (four bands have a resolution of 10 m, six bands have a resolution of 20 m, three bands have a resolution of 60 m). Level-1C data products provide Top-Of-Atmosphere reflectances measured through the MSI as 100 × 100 km<sup>2</sup> ortho-images (tiles) in UTM/WGS84 projection [54]. Level-1C data covers all continental land and sea water up to 20 km from the coast, from June 2015 to the current date [55].

Sentinel-2 data has been selected since this mission provides better performances compared to other public continuous Earth observation missions, in terms of both spatial resolution and revisit period. Landsat 8/9 mission, for example, has a spatial resolution of at most 15 m and a revisit period of 8 days. The choice of satellite imagery data sources, however, is not a conditioning factor for the dataset creation process and other products can be used with few trivial changes in the procedure.

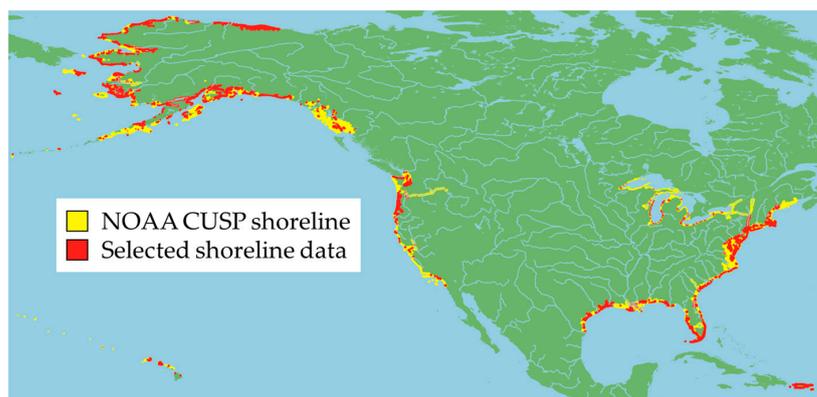
#### 3.1.2. Shoreline Data

CUSP is developed by U.S. NOAA with the aim of providing essential information to manage coastal areas and conduct environmental analyses. This dataset includes all continental U.S. shoreline with portions of Alaska, Hawaii, the U.S. Virgin Islands, Pacific Islands, and Puerto Rico. CUSP provides the mean-high water shoreline, measured through vertical modeling or image interpretation using both water level stations and/or shoreline indicators. All data included in CUSP is verified by contemporary imagery or shoreline from other sources [56]. Another important feature of CUSP is that the shoreline is split into shorter paths and each of them has additional information associated, including the date and type of source data used to measure the shoreline, the type of coast and the horizontal accuracy, which represents the circular error at the 95% confidence level [57]. An analysis of the horizontal accuracy shows that 90% of paths have measurement errors ≤ 10 m, while 99.97% of paths have measurement errors ≤ 20 m. NOAA's CUSP paths have therefore a very high accuracy, comparable and, in most cases, overcoming the resolution of Sentinel 2 imagery. To our knowledge, NOAA's CUSP is the only publicly available source of shoreline data with these features, which are essential to perform the dataset generation procedure proposed in this work. In principle, nothing prevents one from using other sets of shoreline measurements, with the same essential features, i.e., geographic coordinates, date, high accuracy, and possibly the measurement method.

### 3.2. Shoreline Data Preprocessing (Selection and Merging)

A preliminary filtering operation is performed on CUSP data to exclude shoreline that has been extracted from observations prior to the Sentinel-2 mission launch in June 2015. A representation of the shoreline remaining after this preliminary operation is depicted in the map of North America in Figure 1 which shows that locations of useful shorelines are very heterogeneous, spanning most areas of the U.S. coast. This is an advantageous feature since it guarantees a great variability of the satellite images included in the final produced dataset.

Selected shoreline paths that share one terminal point and have the same date are then merged, in order to optimize the satellite images selection procedure described in the following. Statistics about the NOAA CUSP shoreline data and the selected paths are reported in Table 2.



**Figure 1.** Selected CUSP shorelines (those from June 2015 onwards) compared to the complete data.

**Table 2.** Statistics about the NOAA’s CUSP shoreline data.

|                                |            |
|--------------------------------|------------|
| Initial number of paths        | 779,954    |
| Total length of the paths      | 403,707 km |
| Number of selected paths       | 221,331    |
| Total length of selected paths | 107,600 km |
| Number of paths after merging  | 126,938    |

### 3.3. Selection of Satellite Images

In the practical implementation of the procedure, we have obtained the Sentinel-2 Level-1C tiles of our interest, which are  $10,980 \times 10,980$  pixels, using the *Plateforme d’Exploitation des Produits Sentinel* of the *Centre National d’études Spatiales* (PEPS CNES) [58].

Satellite images are selected on the basis of the location and the date of the merged shoreline paths obtained in the former step. It is important here to clarify the issue of *dates and times* of satellite images and on-field measurements used for labeling.

Obviously, the ideal situation is to have perfect simultaneity between the acquisition of the satellite image and the field measurements of the area it takes. It is easy to understand, however, that this situation is unfeasible and never occurs in practice.

Even when in situ measurements are made specifically for image labeling, simultaneity is not achieved in practice (see e.g., [7]). Satellite images, indeed, are taken at intervals of some days (5 days in the case of Sentinel-2), and an image is not always usable, due to the presence of clouds or other causes: hence, usable images have dates that cannot be chosen as desired by the user and can be spaced between them by many days. For this same reason, satellite monitoring is not designed to keep track of changes that occur in a few hours, but of changes over months and years. In general, one must always choose the image *with the date nearer to that of interest*. When labeling a dataset, the date of the image must be as close as possible to that of the in situ measurement.

On the basis of the above consideration, PEPS CNES has been queried according to the following criteria.

- Sentinel tiles must contain the shoreline path.
- Cloud cover of Sentinel tiles must be lesser than 10% (parameter: *cloud\_cover*).
- Sentinel tiles acquisition date must be at most 30 days (parameter: *time\_difference*) before or after the shoreline date.

The time difference between satellite images and NOAA CUSP measurements is exactly known and recorded in the dataset. It never exceeds the parameter *time\_difference*: otherwise, the data sample is not generated. When more than one result is obtained, the tile having the acquisition date closest to the shoreline date is selected.

It is possible to choose different values for the parameters *cloud\_cover* and *time\_difference*. A value of *cloud\_cover* > 10% generates more dataset samples, but it is more likely that they

will be discarded in the following steps (see Section 3.4), due to the presence of clouds. A parameter *time\_difference* < 30 days generates fewer dataset samples, but the shorelines in the obtained images are likely to be negligibly different since the resolution of Sentinel-2 is 10 m / pixel, and therefore only changes of the order of tens of meters are important. In any case, a *time\_difference* < 10 days is not reasonable due to the revisit time of Sentinel-2.

The selection obtained with these constraints has been found to provide a good compromise between the computational resources required to generate the dataset (directly related to the number of selected tiles) and the final size of the dataset, which can grow if more tiles, and hence more shoreline paths, are considered. It is worth highlighting that:

- (1) The quality of each sample generated with this method is assured by later checks, which are also automatic, being based on Sentinel data themselves (see Section 3, in particular, Section 3.4). For example, the presence of clouds in localized areas of the tile is not detrimental.
- (2) Further releasing the constraints (cloud coverage and time difference) do not lead, ultimately, to a significant increase in the dataset's size.

The described procedure has obtained 987 tiles, containing 102,283 shoreline paths (about 20% less than the overall number of paths).

#### 3.4. Extraction of Samples and Labeling

The selected Sentinel-2 tiles are then downloaded and processed singularly to extract the semantically annotated samples. Two preliminary operations are performed before the actual extraction phase.

First, Level-2A products are generated from Level-1C products using the *sen2cor* processor [59]. Level-2A products are composed of (i) a scene classification (SC) mask and (ii) surface reflectance obtained through atmospheric correction [60]. The SC mask assigns one of 12 classes (including water, vegetated and non-vegetated land, and clouds) to each pixel of the tile and is the only data needed in subsequent steps.

Second, the shoreline paths associated with each Sentinel tile are projected to the plane of the UTM/WGS84 zone containing the Sentinel tile. The UTM coordinates of the vertices of the Sentinel tile in the plane of the UTM zone are also known (they can be downloaded from [61]), and thus the pixel coordinates of the shoreline points inside the tile can be computed.

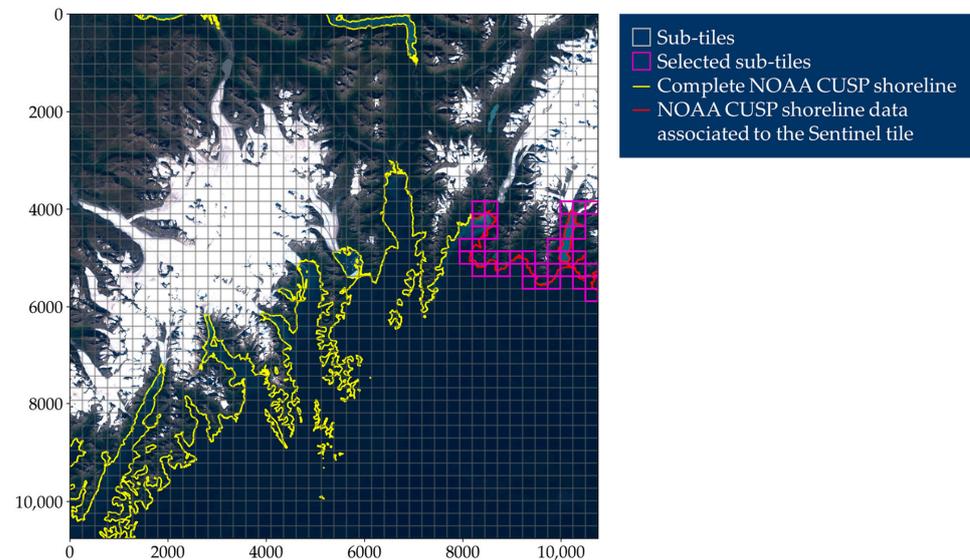
The 10,980 × 10,980 pixels Sentinel tile is then split into sub-tiles of size 256 × 256, among which the samples of the dataset are selected. The sub-tile size has been chosen so that a direct comparison, and a side-to-side utilization, is possible of the obtained dataset and that described in [53], the most recent and numerous for water segmentation in the literature. In Figure 2 an example of a sub-tiling grid is depicted.

The subsequent processing involves only sub-tiles containing shoreline paths, as shown in Figure 2. The basic task is to create, for each sub-tile, a binary segmentation map based on NOAA CUSP shoreline paths. This operation is of some complexity and must be illustrated in detail.

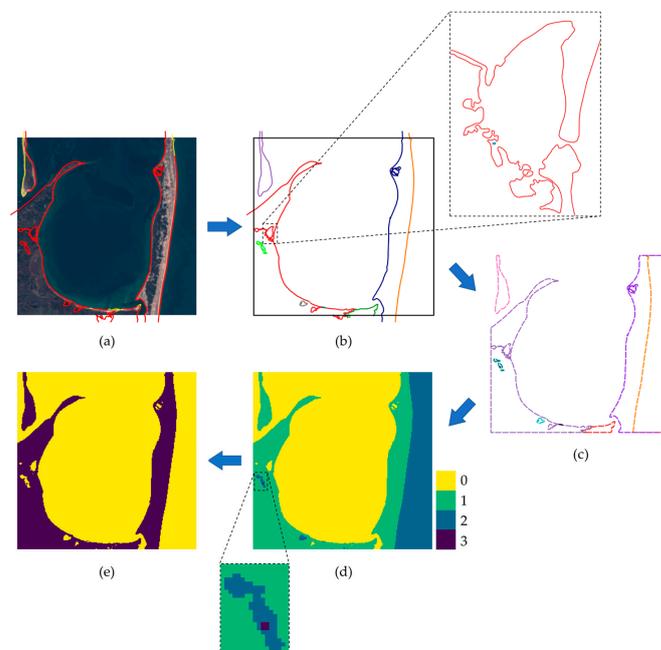
First of all, we specify that, for any sub-tile, all shoreline paths partially or completely contained in it are considered, *independent of their date*. A strict constraint on the dates of all the used shoreline segments leads to discarding many sub-tiles, due to the short shoreline segments with dates too different from that of the tile. Instead, completing the shoreline including also short segments measured in different dates allows the construction of a dataset with much more sample, without compromising meaningfully the quality of the shoreline data. Besides, the date of each shoreline path is supplied in the dataset, so that samples can be later selected, if deemed useful, according to arbitrary constraints on the time difference between the Sentinel date and the CUSP shoreline dates.

The process used for generating the binary segmentation mask of sub-tiles from CUSP shoreline paths is depicted in Figure 3. As a first step, shoreline paths completely or partially contained in the sub-tile are selected (Figure 3a). The second step is to merge contacting paths: merged paths, depicted with different colors in Figure 3b, can be closed

(e.g., the light green path and the gray path) or open (e.g., the orange, blue and red paths). If open paths have an end inside the sub-tile, the sub-tile is discarded; otherwise, paths are clipped using the sub-tile borders as the clipping window. Only closed polygons are obtained after this operation, as shown in Figure 3c. To obtain a binary mask, polygons are filled with ones and summed, producing the matrix in Figure 3d; finally, a binary map (Figure 3e) is obtained by selecting the even and odds elements of the matrix in Figure 3d.

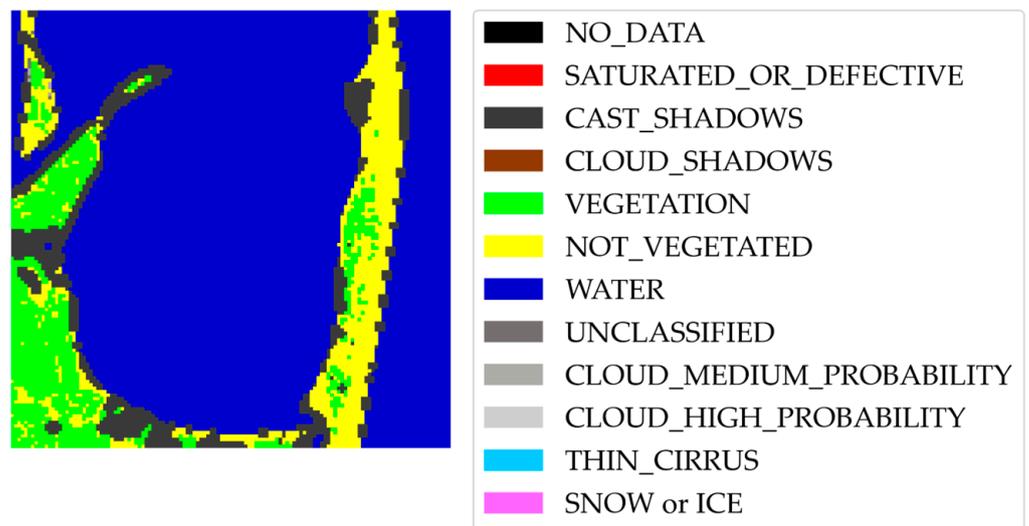


**Figure 2.** Example of Sentinel-2 Level-1C tile (true color image) split into sub-tiles of size  $256 \times 256$ . Only sub-tiles containing shoreline based on measurements performed in dates within 30 days from the Sentinel tile's acquisition date are analyzed.



**Figure 3.** Steps for generating the binary segmentation mask for a sample of the dataset. (a) Selection of the shoreline paths inside the sub-tile. Paths with dates compatible with the Sentinel-2 tile's date are depicted in red and the other paths in yellow. (b) Merging of contacting paths. Distinct merged paths are depicted using different colors, while the sub-tile border is in black. (c) Clipping of merged paths using the sub-tile border as the clipping window. After this stage, closed polygons are obtained. (d) Starting from a zero-filled matrix of the sub-tile, ones are added in the regions defined by the polygons. (e) A binary map is obtained by classifying the pixels of matrix (d) into even and odds.

The segmentation label of the sub-tile is created by assigning water and land categories to the two classes of the mask in Figure 3e based on the previously computed Level-2A SC layer, depicted in Figure 4 for the case considered in Figure 3. The class in the mask containing more Level-2A water pixels is categorized as water, while the class containing more non-water Level-2A pixels is categorized as land. The SC layer is also used to evaluate the correctness of the label. In particular, the sub-tile is discarded if the water class and land class contain less than 80% Level-2A water pixels and non-water pixels, respectively.

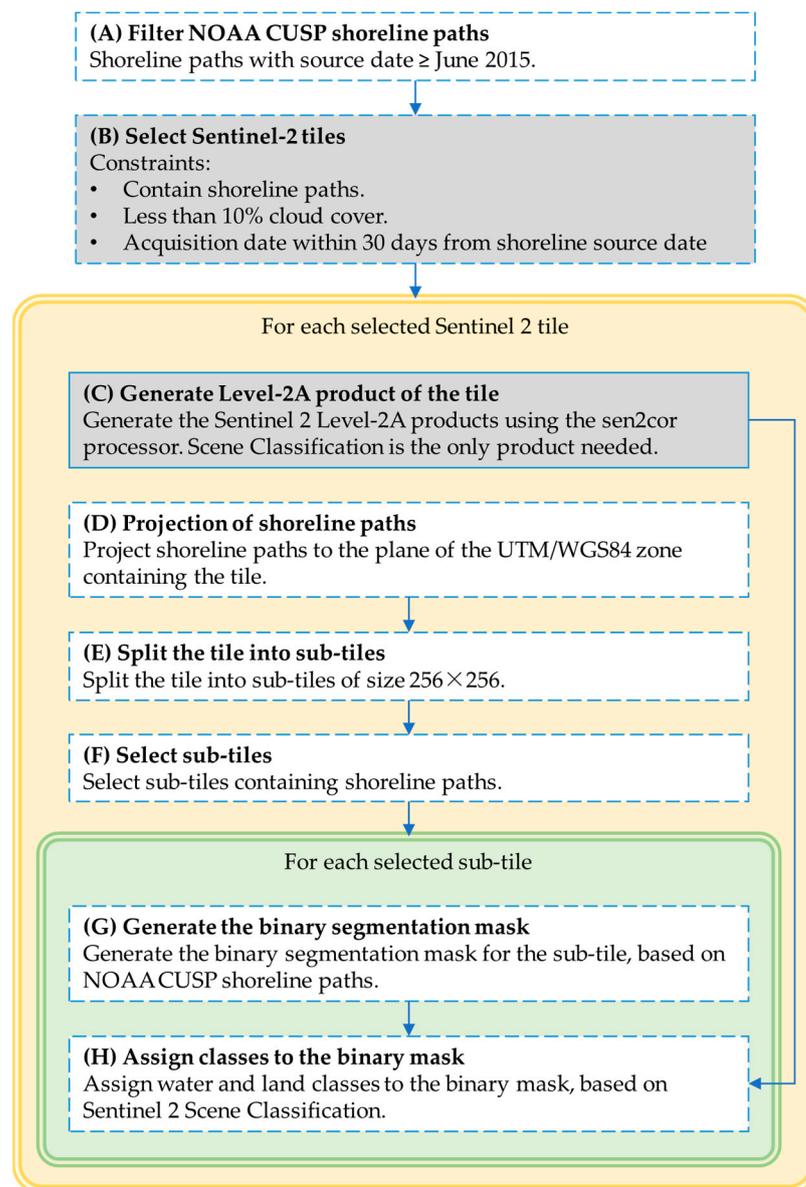


**Figure 4.** Sentinel-2 Level-2A scene classification (SC) mask of the sub-tile in Figure 3. For the sake of clarity, the legend includes all the 12 classes provided by Level-2A SC, although only some of them are identified in this case.

### 3.5. Overview of the Dataset Generation Procedure

For the sake of clarity, a flowchart of the previously described dataset generation method is depicted in Figure 5. The proposed method is an original solution, and each step except one (marked with a solid line) has been designed and implemented purposely for this work. The flowchart highlights the novelty of this method compared to others reported in the literature and reviewed in Section 2. In these works, sea/land labeling is fundamentally based on the human interpretation of satellite imagery, while the proposed method uses shoreline measurements from NOAA CUSP as a source for automatic labeling.

In the flowchart in Figure 5, operations with a gray background, namely B and C, are specific to Sentinel 2 and need major refactoring if other imagery sources are used. All the other operations require instead only trivial changes. The general logic and overall procedure for dataset generation are, however, the same even if other imagery sources are used. Furthermore, the changes required for operations B and C are not substantial. In particular, for operation B, the same constraints for selecting satellite images must be used to query the appropriate satellite imagery platform (PEPS CNES is used in this work for Sentinel images). For operation C, the only required output is a low-accuracy sea/land classification of the satellite image, used later in operation H, and this can be easily obtained e.g., by computing the NDWI.



**Figure 5.** Flowchart of the proposed method. Details about each operation are provided in the previous subsections. Operations performed with novel methods, proposed in the paper, are marked with a dashed blue line. Operations performed with known methods are marked with a solid blue line. Operations with gray backgrounds are those specific for Sentinel 2 imagery.

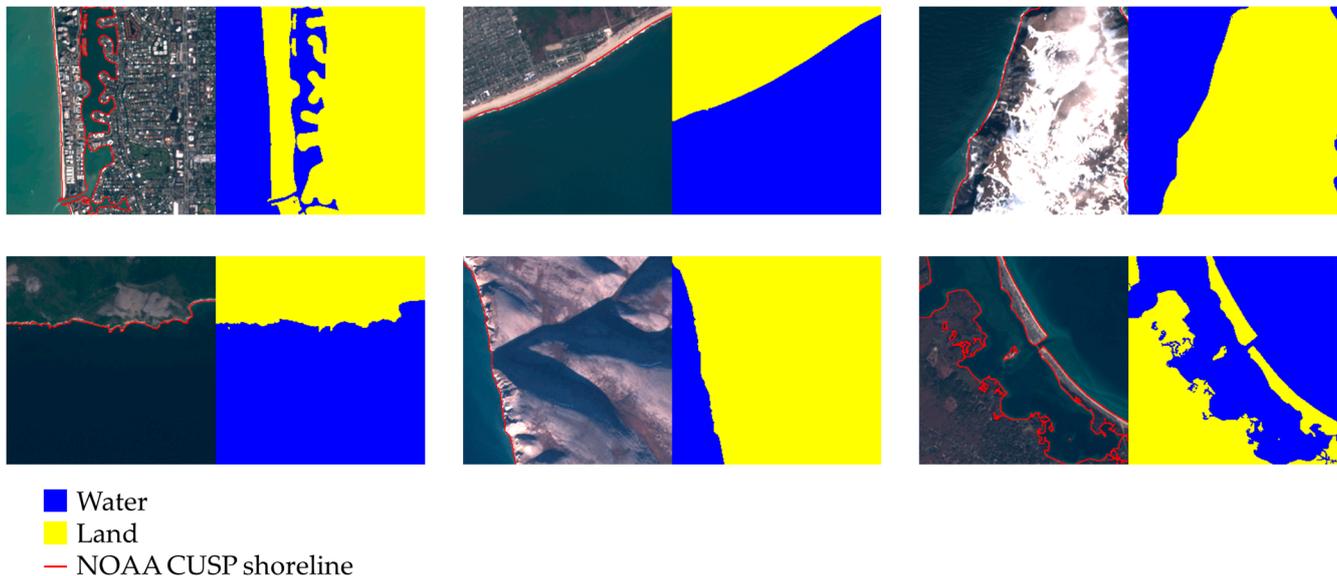
#### 4. Results and Discussion

The annotated dataset generated using the proposed method counts 4334 samples of size  $256 \times 256$  containing both water and land pixels. The dataset has been built retrieving all 13 Sentinel-2 MSI bands, which are therefore all present in each sample. The resolution of Sentinel-2 images is different for the various bands, the best being 10 m per pixel. The images in all 13 bands have been linearly rescaled to a uniform 10 m per pixel spatial resolution, a standard operation that allows one to have all the images in a single 3-d array. Each sample is provided with the water/land segmentation label and with the following additional information.

- Level-2A SC mask.
- Shoreline paths are used for labeling, each with its measurement date.
- PEPS CNES identifier of the Sentinel-2 Level-1C tile.
- Acquisition date of the Sentinel-2 Level-1C tile.

- Pixels offset of the sub-tile in the complete Sentinel-2 image.

Some examples taken from the generated dataset are depicted in Figure 6. It is possible to appreciate the accuracy of the labeling, especially in the two cases of complicated water edges.



**Figure 6.** Examples of semantically annotated Sentinel-2 satellite images (true color image on the left) and labels (on the right) contained in the proposed dataset.

In the next subsections, the quality of the dataset is assessed both visually, by comparing images of a random subset of the dataset with their corresponding labels, and from a functional point of view, by training and testing a deep learning model using SNOWED.

#### 4.1. Dataset Visual Quality Assessment

In premise, it is important to remember that *any* dataset is prone to including inconsistent labels. In datasets that are manually labeled by subject matter experts there is room for human mistakes and subjective interpretations; in automatically labeled datasets problems may arise from intrinsic imperfections in the labeling algorithm. The problem of inconsistent labels is negligible only in synthetically generated datasets, which, in contrast, are prone to providing samples not similar enough to the actual samples with which the model must work. Therefore, assessing the quality of a dataset, and providing methods to improve it, can be considered a good metrological practice, always advisable.

The automatic method presented here to construct the dataset, together with its clear advantages, has an intrinsic drawback, that occasionally produces samples with incomplete labels. The problem arises from the fact that, in general, there is no guarantee that *the set of measured shorelines includes all the water edges in each sample*. In the case of SNOWED, a Sentinel-2 sub-tile may include water edges that have not been measured and included in CUSP NOAA. This problem can be avoided only by using a (hypothetical) collection of shorelines that is guaranteed to include *all* the water edges present in a large enough geographic region. This is not the case with NOAA CUSP.

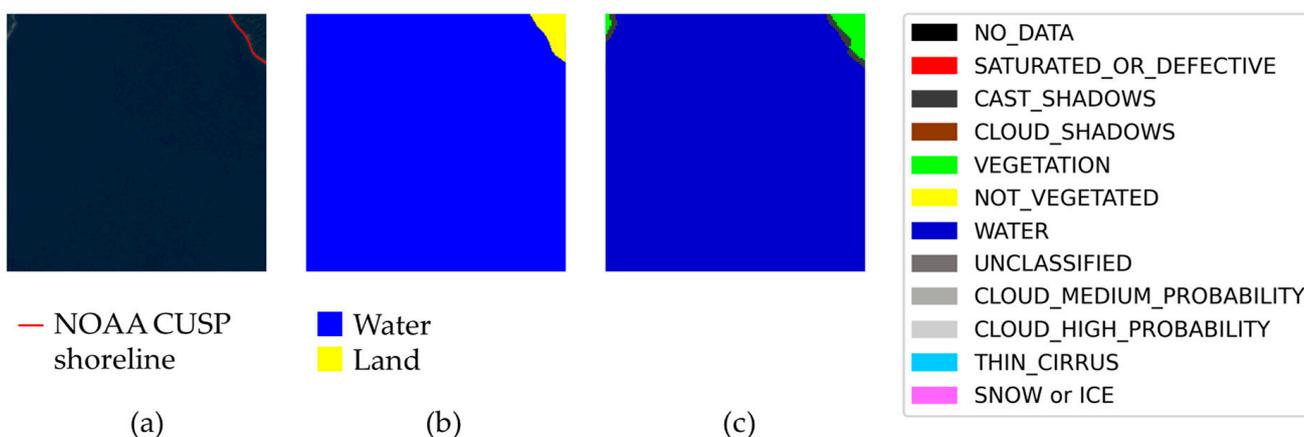
Because of the considerations above, we provide here a procedure to check, assess and improve the dataset quality.

Any single sample of the dataset can be checked by inspecting three images provided in it, i.e., the TCI, the label, and the Sentinel-2 scene classification, as shown in Figure 7. The TCI and the label are visually compared, and the Sentinel-2 SC is used as a guide. It is important to remember that the latter image may only serve as a guide for a human, and not for an automatic check: the scene classification is, indeed, not very accurate, and in some cases misclassifies regions of the satellite image.

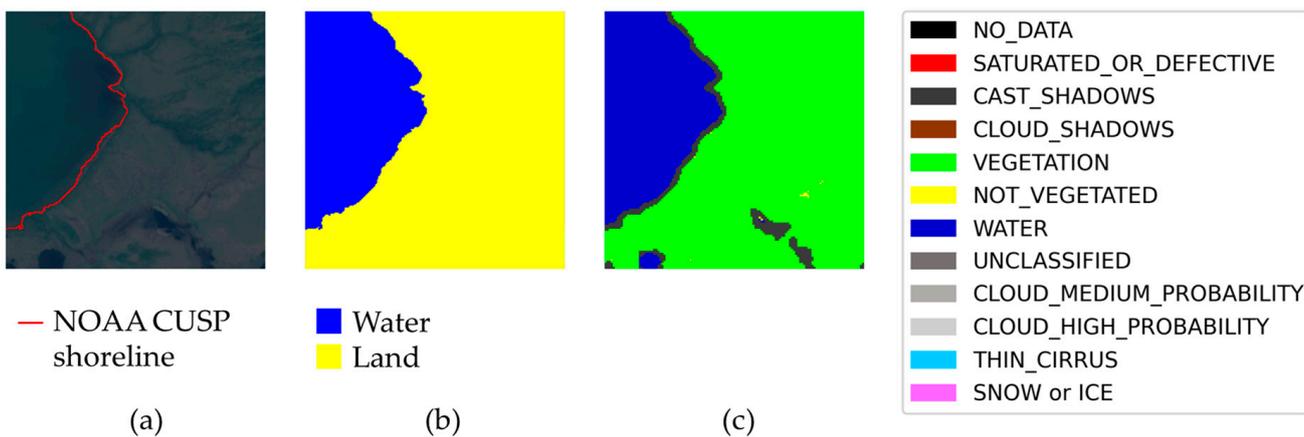
In Figure 7 it is clear that, in the upper left corner, there is a small water edge, and therefore a small portion of land, not included in the label. This water edge was simply not present in NOAA CUSP and is of a length comparable with that of the labeled water edge in the sample. This sample, therefore, is considered “bad”.

In Figure 8, instead, there is a sample that we consider “suspect”. In this sample, the main shoreline is clear and labeled, but there is a small region which is classified as water by Sentinel-2 SC and as land according NOAA CUSP. It is not obvious if the water edge is real or not—a further check with an independent source should be made, e.g., using commercial satellite imagery with very high resolution—and the (possibly) missing edge is much shorter than the labeled one.

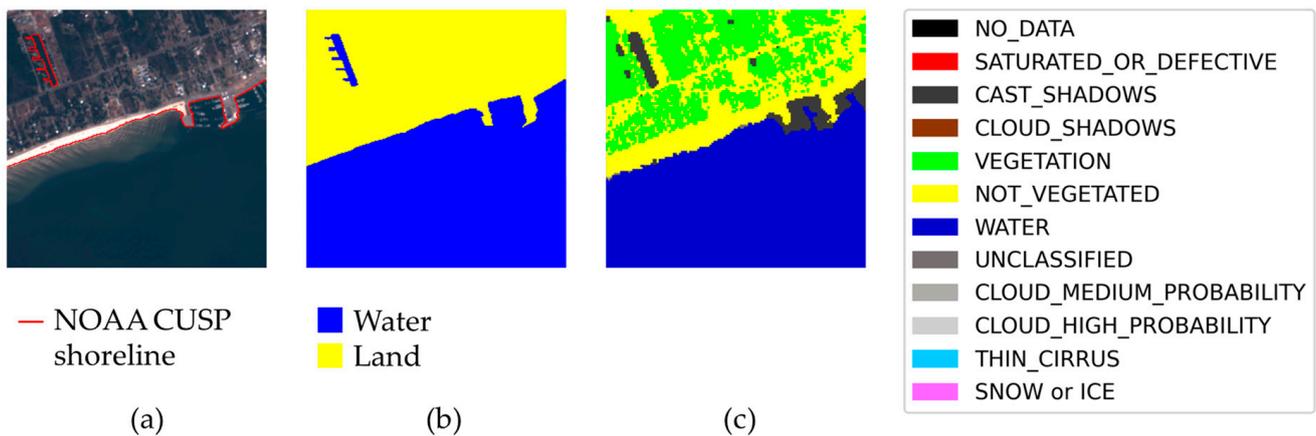
We want to highlight that, together with “bad” and “suspect” samples, the dataset has many samples that are “particularly good”, in the sense that the label includes elaborated shorelines, difficult to recognize by a human, and that the label is also completely ignored by Sentinel-2 SC. An example is in Figure 9.



**Figure 7.** Examples of “bad” samples with a clearly incomplete label, found in the process of quality assessment of the dataset. (a) True color image. (b) Sea/land segmentation. (c) Sentinel-2 Level-2A scene classification. There is a portion of land not included in the label, in the upper left corner.



**Figure 8.** Example of a “suspect” sample, with a possibly incomplete label. There is a region that is water according to Sentinel-2 scene interpretation. The main shoreline in the sample is labeled. (a) True color image. (b) Sea/land segmentation. (c) Sentinel-2 Level-2A scene classification.



**Figure 9.** Example of a “particularly good” sample, with an elaborated water edge that is ignored by Sentinel-2 SC. (a) True color image. (b) Sea/land segmentation. (c) Sentinel-2 Level-2A scene classification.

We have assessed the dataset quality by checking a randomly selected subset of  $n = 200$  samples, out of  $N = 4334$  sample total. In the selected samples, we found five “bad” samples, with clearly incomplete labels, and 30 “suspect” samples, with possibly incomplete labels and ambiguous interpretation. The point estimate of the fraction of “bad” and “suspect” samples in a dataset (a conservative estimate of the fraction of improvable samples) is therefore:

$$\hat{p} = \frac{35}{200} = 17.5\%$$

and the 95% confidence interval for this fraction is, approximately:

$$p_{95\%} = \hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n} \frac{N-n}{N-1}} = (17.5 \pm 5.1)\%$$

where the Gaussian approximation of the hypergeometric distribution has been applied, taking into account the term  $(N-n)/(N-1)$  to correctly account for the sampling without replacement in the acceptance sampling procedure [62,63].

The samples in this fraction can be further processed, by humans or by an algorithm using a different source of water edge data, to improve the labeling. It can also be discarded, even if this choice does not seem appropriate for the “suspect” samples, whose labels always include the “main” shorelines in the image.

In the next subsection, the dataset is used “as is”, without discarding or correcting neither the bad samples nor the suspect ones found in the assessment process. It is shown that the dataset trains quite effectively a simple neural model for shoreline detection.

#### 4.2. Example Application of the Dataset

The SNOWED dataset has been employed to train a standard U-Net segmentation model [31]. The dataset has been shuffled and then split into a training and a validation subset, corresponding to 80% and 20% of the samples respectively. Afterwards, the U-Net neural network has been trained for 30 epochs, using the Adam optimizer [64]. Cross-entropy has been used as a loss function in the optimization, while mean intersection over union (IoU) is the metric to evaluate the performance of the neural model.

The training process is depicted in Figure 10. The final mean IoU for the validation set, obtained after completing the training phase, is 0.936. In Figure 11, the sea/land segmentation masks produced by the trained U-Net model for the first four samples of the validation set are depicted. The visual inspection of these results shows that an accurate sea/water segmentation is obtained, even if we used the standard U-Net model without any further optimization.

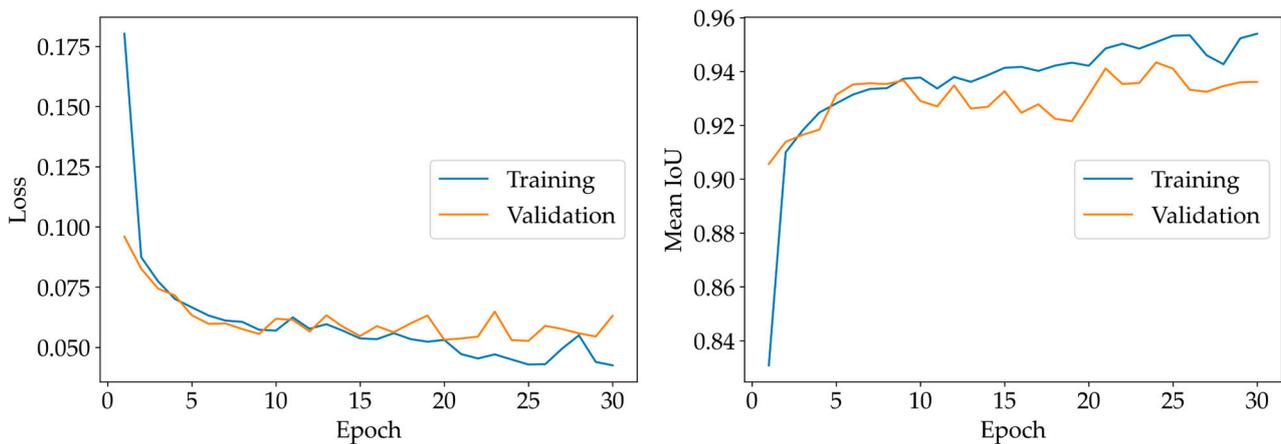


Figure 10. Loss function and mean IoU versus number of training epochs.

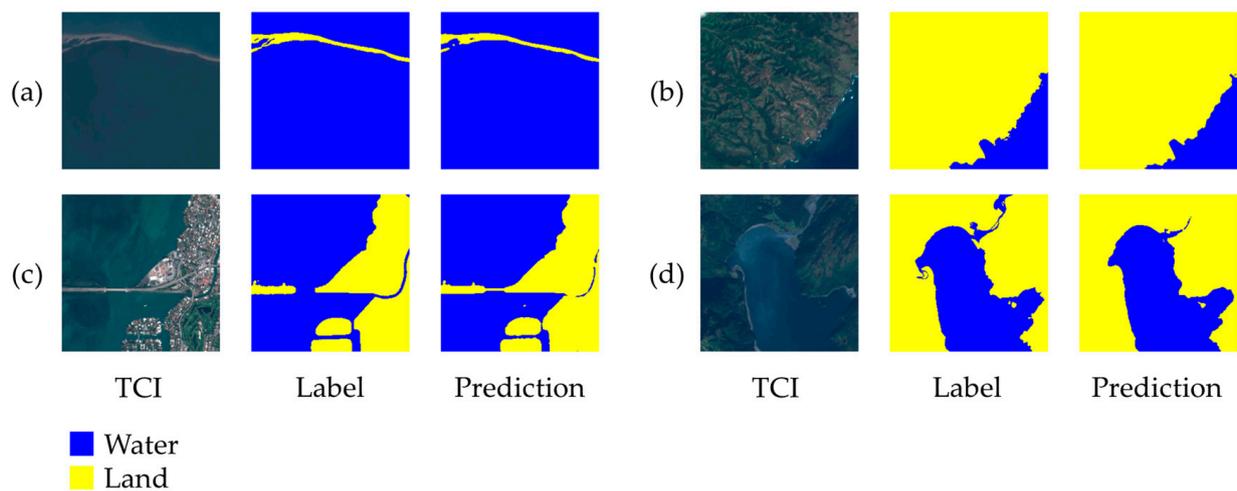


Figure 11. (a–d) Sea/land segmentation results obtained with the trained U-Net model for the first four samples of the validation set.

## 5. Conclusions

Measuring boundaries between land and water is important for understanding and managing environmental phenomena like erosion, accretion, sea level rise, etc. Measurements from satellite imagery are particularly useful to provide consistent information over large areas and long periods of time (even if with limited spatial resolution). There is not a single best method to measure shorelines using satellite data, but artificial intelligence techniques are also acquiring more and more importance in this field. Therefore, recent research is devoted to the construction of datasets of satellite images with shoreline labels, customarily obtained with human work of image interpretation and annotation. Constructing datasets with human intervention has obvious costs and drawbacks, which are especially meaningful considering that a dataset of labeled images of a given satellite cannot be used to work with images of other satellites.

Based on these considerations, we have focused on the task of constructing a labeled dataset of satellite images for shoreline detection *without any human intervention*. The algorithm uses NOAA CUSP shoreline data to properly select and annotate satellite images. By annotating Sentinel-2 Level-1C images, the algorithm has constructed the SNOWED dataset, which can be used alongside the very recent SWED dataset. With minimal adjustments, the algorithm can be used to construct datasets for different satellites.

The concept and results presented in this work show that it is possible, in general, to construct readily a dataset of labeled satellite images, if a set of in situ measurements

with geographic and temporal data are available. Therefore, satellite monitoring and measurements can receive a great boost by increasing the public availability of measurement data coming from accurate ground surveys.

**Author Contributions:** Conceptualization, M.S. (Marco Scarpetta), M.S. (Maurizio Spadavecchia), P.A. and N.G.; Methodology, G.A., M.S. (Marco Scarpetta), M.S. (Maurizio Spadavecchia), P.A. and N.G.; Software, M.S. (Marco Scarpetta) and P.A.; Validation, M.S. (Marco Scarpetta), M.S. (Maurizio Spadavecchia) and N.G.; Formal analysis, M.S. (Marco Scarpetta), M.S. (Maurizio Spadavecchia) and N.G.; Investigation, G.A., M.S. (Marco Scarpetta), M.S. (Maurizio Spadavecchia), P.A. and N.G.; Resources, M.S. (Maurizio Spadavecchia) and N.G.; Data curation, M.S. (Marco Scarpetta) and P.A.; Writing—original draft, M.S. (Marco Scarpetta), M.S. (Maurizio Spadavecchia) and N.G.; Writing—review & editing, G.A., M.S. (Marco Scarpetta), M.S. (Maurizio Spadavecchia), P.A. and N.G.; Visualization, G.A., M.S. (Marco Scarpetta), M.S. (Maurizio Spadavecchia) and N.G.; Supervision, M.S. (Maurizio Spadavecchia) N.G.; Project administration, G.A., M.S. (Maurizio Spadavecchia) and N.G.; Funding acquisition, G.A., N.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Polytechnic University of Bari and by research project PON-MITIGO (ARS01\_00964).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are openly available in Zenodo at <https://doi.org/10.5281/Zenodo.7871636>, reference number [48].

**Acknowledgments:** The authors wish to thank Vito Ivano D’Alessandro and Luisa De Palma for their useful hints and discussions in the initial stage of the work.

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

## References

1. Martínez, M.L.; Intralawan, A.; Vázquez, G.; Pérez-Maqueo, O.; Sutton, P.; Landgrave, R. The Coasts of Our World: Ecological, Economic and Social Importance. *Ecol. Econ.* **2007**, *63*, 254–272. [CrossRef]
2. Halpern, B.S.; Frazier, M.; Afflerbach, J.; Lowndes, J.S.; Micheli, F.; O’Hara, C.; Scarborough, C.; Selkoe, K.A. Recent Pace of Change in Human Impact on the World’s Ocean. *Sci. Rep.* **2019**, *9*, 11609. [CrossRef] [PubMed]
3. Adamo, F.; Andria, G.; Cavone, G.; De Capua, C.; Lanzolla, A.M.L.; Morello, R.; Spadavecchia, M. Estimation of Ship Emissions in the Port of Taranto. *Measurement* **2014**, *47*, 982–988. [CrossRef]
4. Cotecchia, F.; Vitone, C.; Sollecito, F.; Mali, M.; Miccoli, D.; Petti, R.; Milella, D.; Ruggieri, G.; Bottiglieri, O.; Santaloia, F.; et al. A Geo-Chemo-Mechanical Study of a Highly Polluted Marine System (Taranto, Italy) for the Enhancement of the Conceptual Site Model. *Sci. Rep.* **2021**, *11*, 4017. [CrossRef]
5. Tiwari, M.; Rathod, T.D.; Ajmal, P.Y.; Bhangare, R.C.; Sahu, S.K. Distribution and Characterization of Microplastics in Beach Sand from Three Different Indian Coastal Environments. *Mar. Pollut. Bull.* **2019**, *140*, 262–273. [CrossRef] [PubMed]
6. Vedolin, M.C.; Teophilo, C.Y.S.; Turra, A.; Figueira, R.C.L. Spatial Variability in the Concentrations of Metals in Beached Microplastics. *Mar. Pollut. Bull.* **2018**, *129*, 487–493. [CrossRef]
7. Traganos, D.; Aggarwal, B.; Poursanidis, D.; Topouzelis, K.; Chrysoulakis, N.; Reinartz, P. Towards Global-Scale Seagrass Mapping and Monitoring Using Sentinel-2 on Google Earth Engine: The Case Study of the Aegean and Ionian Seas. *Remote Sens.* **2018**, *10*, 1227. [CrossRef]
8. Scarpetta, M.; Affuso, P.; De Virgilio, M.; Spadavecchia, M.; Andria, G.; Giaquinto, N. Monitoring of Seagrass Meadows Using Satellite Images and U-Net Convolutional Neural Network. In Proceedings of the 2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Ottawa, ON, Canada, 16–19 May 2022; pp. 1–6.
9. Adamo, F.; Attivissimo, F.; Carducci, C.G.C.; Lanzolla, A.M.L. A Smart Sensor Network for Sea Water Quality Monitoring. *IEEE Sens. J.* **2015**, *15*, 2514–2522. [CrossRef]
10. Attivissimo, F.; Carducci, C.G.C.; Lanzolla, A.M.L.; Massaro, A.; Vadrucci, M.R. A Portable Optical Sensor for Sea Quality Monitoring. *IEEE Sens. J.* **2015**, *15*, 146–153. [CrossRef]
11. Lu, J.; Wu, J.; Zhang, C.; Zhang, Y.; Lin, Y.; Luo, Y. Occurrence, Distribution, and Ecological-Health Risks of Selected Antibiotics in Coastal Waters along the Coastline of China. *Sci. Total Environ.* **2018**, *644*, 1469–1476. [CrossRef]
12. Zhang, R.; Pei, J.; Zhang, R.; Wang, S.; Zeng, W.; Huang, D.; Wang, Y.; Zhang, Y.; Wang, Y.; Yu, K. Occurrence and Distribution of Antibiotics in Mariculture Farms, Estuaries and the Coast of the Beibu Gulf, China: Bioconcentration and Diet Safety of Seafood. *Ecotoxicol. Environ. Saf.* **2018**, *154*, 27–35. [CrossRef]

13. Kaku, K. Satellite Remote Sensing for Disaster Management Support: A Holistic and Staged Approach Based on Case Studies in Sentinel Asia. *Int. J. Disaster Risk Reduct.* **2019**, *33*, 417–432. [[CrossRef](#)]
14. Sòria-Perpinyà, X.; Vicente, E.; Urrego, P.; Pereira-Sandoval, M.; Tenjo, C.; Ruíz-Verdú, A.; Delegido, J.; Soria, J.M.; Peña, R.; Moreno, J. Validation of Water Quality Monitoring Algorithms for Sentinel-2 and Sentinel-3 in Mediterranean Inland Waters with In Situ Reflectance Data. *Water* **2021**, *13*, 686. [[CrossRef](#)]
15. Angelini, M.G.; Costantino, D.; Di Nisio, A. ASTER Image for Environmental Monitoring Change Detection and Thermal Map. In Proceedings of the 2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Turin, Italy, 22–25 May 2017; pp. 1–6.
16. Spinosa, A.; Ziemba, A.; Saponieri, A.; Damiani, L.; El Serafy, G. Remote Sensing-Based Automatic Detection of Shoreline Position: A Case Study in Apulia Region. *J. Mar. Sci. Eng.* **2021**, *9*, 575. [[CrossRef](#)]
17. Scarpetta, M.; Spadavecchia, M.; Andria, G.; Ragolia, M.A.; Giaquinto, N. Simultaneous Measurement of Heartbeat Intervals and Respiratory Signal Using a Smartphone. In Proceedings of the 2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Lausanne, Switzerland, 23–25 June 2021; pp. 1–5.
18. Abdelhady, H.U.; Troy, C.D.; Habib, A.; Manish, R. A Simple, Fully Automated Shoreline Detection Algorithm for High-Resolution Multi-Spectral Imagery. *Remote Sens.* **2022**, *14*, 557. [[CrossRef](#)]
19. Sekar, C.S.; Kankara, R.S.; Kalaivanan, P. Pixel-Based Classification Techniques for Automated Shoreline Extraction on Open Sandy Coast Using Different Optical Satellite Images. *Arab. J. Geosci.* **2022**, *15*, 939. [[CrossRef](#)]
20. Ragolia, M.A.; Andria, G.; Attivissimo, F.; Nisio, A.D.; Maria Lucia Lanzolla, A.; Spadavecchia, M.; Larizza, P.; Brunetti, G. Performance Analysis of an Electromagnetic Tracking System for Surgical Navigation. In Proceedings of the 2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Istanbul, Turkey, 26–28 June 2019; pp. 1–6.
21. De Palma, L.; Scarpetta, M.; Spadavecchia, M. Characterization of Heart Rate Estimation Using Piezoelectric Plethysmography in Time- and Frequency-Domain. In Proceedings of the 2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Bari, Italy, 1 June–1 July 2020; pp. 1–6.
22. Alcaras, E.; Falchi, U.; Parente, C.; Vallario, A. Accuracy Evaluation for Coastline Extraction from Pléiades Imagery Based on NDWI and IHS Pan-Sharpener Application. *Appl. Geomat.* **2022**. [[CrossRef](#)]
23. Dai, C.; Howat, I.M.; Larour, E.; Husby, E. Coastline Extraction from Repeat High Resolution Satellite Imagery. *Remote Sens. Environ.* **2019**, *229*, 260–270. [[CrossRef](#)]
24. Al Ruheili, A.M.; Boluwade, A. Quantifying Coastal Shoreline Erosion Due to Climatic Extremes Using Remote-Sensed Estimates from Sentinel-2A Data. *Environ. Process.* **2021**, *8*, 1121–1140. [[CrossRef](#)]
25. Sunny, D.S.; Islam, K.M.A.; Mullick, M.R.A.; Ellis, J.T. Performance Study of Imageries from MODIS, Landsat 8 and Sentinel-2 on Measuring Shoreline Change at a Regional Scale. *Remote Sens. Appl. Soc. Environ.* **2022**, *28*, 100816. [[CrossRef](#)]
26. Sánchez-García, E.; Palomar-Vázquez, J.M.; Pardo-Pascual, J.E.; Almonacid-Caballer, J.; Cabezas-Rabadán, C.; Gómez-Pujol, L. An Efficient Protocol for Accurate and Massive Shoreline Definition from Mid-Resolution Satellite Imagery. *Coast. Eng.* **2020**, *160*, 103732. [[CrossRef](#)]
27. Cabezas-Rabadán, C.; Pardo-Pascual, J.E.; Palomar-Vázquez, J.; Fernández-Sarría, A. Characterizing Beach Changes Using High-Frequency Sentinel-2 Derived Shorelines on the Valencian Coast (Spanish Mediterranean). *Sci. Total Environ.* **2019**, *691*, 216–231. [[CrossRef](#)] [[PubMed](#)]
28. Novellino, A.; Engwell, S.L.; Grebby, S.; Day, S.; Cassidy, M.; Madden-Nadeau, A.; Watt, S.; Pyle, D.; Abdurrachman, M.; Edo Marshal Nurshal, M.; et al. Mapping Recent Shoreline Changes Spanning the Lateral Collapse of Anak Krakatau Volcano, Indonesia. *Appl. Sci.* **2020**, *10*, 536. [[CrossRef](#)]
29. Guo, Z.; Wu, L.; Huang, Y.; Guo, Z.; Zhao, J.; Li, N. Water-Body Segmentation for SAR Images: Past, Current, and Future. *Remote Sens.* **2022**, *14*, 1752. [[CrossRef](#)]
30. Tsiakos, C.-A.D.; Chalkias, C. Use of Machine Learning and Remote Sensing Techniques for Shoreline Monitoring: A Review of Recent Literature. *Appl. Sci.* **2023**, *13*, 3268. [[CrossRef](#)]
31. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Part III, Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
32. Baumhoer, C.A.; Dietz, A.J.; Kneisel, C.; Kuenzer, C. Automated Extraction of Antarctic Glacier and Ice Shelf Fronts from Sentinel-1 Imagery Using Deep Learning. *Remote Sens.* **2019**, *11*, 2529. [[CrossRef](#)]
33. Heidler, K.; Mou, L.; Baumhoer, C.; Dietz, A.; Zhu, X.X. HED-UNet: Combined Segmentation and Edge Detection for Monitoring the Antarctic Coastline. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4300514. [[CrossRef](#)]
34. Zhang, S.; Xu, Q.; Wang, H.; Kang, Y.; Li, X. Automatic Waterline Extraction and Topographic Mapping of Tidal Flats From SAR Images Based on Deep Learning. *Geophys. Res. Lett.* **2022**, *49*, e2021GL096007. [[CrossRef](#)]
35. Aghdami-Nia, M.; Shah-Hosseini, R.; Rostami, A.; Homayouni, S. Automatic Coastline Extraction through Enhanced Sea-Land Segmentation by Modifying Standard U-Net. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *109*, 102785. [[CrossRef](#)]
36. Shamsolmoali, P.; Zareapoor, M.; Wang, R.; Zhou, H.; Yang, J. A Novel Deep Structure U-Net for Sea-Land Segmentation in Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3219–3232. [[CrossRef](#)]
37. Li, R.; Liu, W.; Yang, L.; Sun, S.; Hu, W.; Zhang, F.; Li, W. DeepUNet: A Deep Fully Convolutional Network for Pixel-Level Sea-Land Segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3954–3962. [[CrossRef](#)]

38. Cheng, D.; Meng, G.; Xiang, S.; Pan, C. FusionNet: Edge Aware Deep Convolutional Networks for Semantic Segmentation of Remote Sensing Harbor Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 5769–5783. [[CrossRef](#)]
39. Cheng, D.; Meng, G.; Cheng, G.; Pan, C. SeNet: Structured Edge Network for Sea–Land Segmentation. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 247–251. [[CrossRef](#)]
40. Cui, B.; Jing, W.; Huang, L.; Li, Z.; Lu, Y. SANet: A Sea–Land Segmentation Network Via Adaptive Multiscale Feature Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 116–126. [[CrossRef](#)]
41. Dang, K.B.; Dang, V.B.; Ngo, V.L.; Vu, K.C.; Nguyen, H.; Nguyen, D.A.; Nguyen, T.D.L.; Pham, T.P.N.; Giang, T.L.; Nguyen, H.D.; et al. Application of Deep Learning Models to Detect Coastlines and Shorelines. *J. Environ. Manag.* **2022**, *320*, 115732. [[CrossRef](#)] [[PubMed](#)]
42. Tajima, Y.; Wu, L.; Watanabe, K. Development of a Shoreline Detection Method Using an Artificial Neural Network Based on Satellite SAR Imagery. *Remote Sens.* **2021**, *13*, 2254. [[CrossRef](#)]
43. Jing, W.; Cui, B.; Lu, Y.; Huang, L. BS-Net: Using Joint-Learning Boundary and Segmentation Network for Coastline Extraction from Remote Sensing Images. *Remote Sens. Lett.* **2021**, *12*, 1260–1268. [[CrossRef](#)]
44. Scarpetta, M.; Spadavecchia, M.; Adamo, F.; Ragolia, M.A.; Giaquinto, N. Detection and Characterization of Multiple Discontinuities in Cables with Time-Domain Reflectometry and Convolutional Neural Networks. *Sensors* **2021**, *21*, 8032. [[CrossRef](#)]
45. Frolov, V.; Faizov, B.; Shakhuro, V.; Sanzharov, V.; Konushin, A.; Galaktionov, V.; Voloboy, A. Image Synthesis Pipeline for CNN-Based Sensing Systems. *Sensors* **2022**, *22*, 2080. [[CrossRef](#)]
46. Scarpetta, M.; Spadavecchia, M.; Andria, G.; Ragolia, M.A.; Giaquinto, N. Analysis of TDR Signals with Convolutional Neural Networks. In Proceedings of the 2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Virtual, 17–20 May 2021; pp. 1–6.
47. Scarpetta, M.; Spadavecchia, M.; D’Alessandro, V.I.; Palma, L.D.; Giaquinto, N. A New Dataset of Satellite Images for Deep Learning-Based Coastline Measurement. In Proceedings of the 2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRINE), Rome, Italy, 26–28 October 2022; pp. 635–640.
48. Andria, G.; Scarpetta, M.; Spadavecchia, M.; Affuso, P.; Giaquinto, N. Sentinel2-NOAA Water Edges Dataset (SNOWED). Available online: <https://doi.org/10.5281/Zenodo.7871636> (accessed on 27 April 2023).
49. QueryPlanet Water Segmentation Data Set. Available online: <http://queryplanet.sentinel-hub.com/index.html?prefix=#waterdata> (accessed on 28 June 2022).
50. Mcfeeters, S.K. The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [[CrossRef](#)]
51. Yang, T.; Jiang, S.; Hong, Z.; Zhang, Y.; Han, Y.; Zhou, R.; Wang, J.; Yang, S.; Tong, X.; Kuc, T. Sea-Land Segmentation Using Deep Learning Techniques for Landsat-8 OLI Imagery. *Mar. Geod.* **2020**, *43*, 105–133. [[CrossRef](#)]
52. Erdem, F.; Bayram, B.; Bakirman, T.; Bayrak, O.C.; Akpınar, B. An Ensemble Deep Learning Based Shoreline Segmentation Approach (WaterNet) from Landsat 8 OLI Images. *Adv. Space Res.* **2021**, *67*, 964–974. [[CrossRef](#)]
53. Seale, C.; Redfern, T.; Chatfield, P.; Luo, C.; Dempsey, K. Coastline Detection in Satellite Imagery: A Deep Learning Approach on New Benchmark Data. *Remote Sens. Environ.* **2022**, *278*, 113044. [[CrossRef](#)]
54. Snyder, J.P. *Map Projections—A Working Manual*; US Government Printing Office: Washington, DC, USA, 1987; Volume 1395.
55. Sentinel-2—Missions—Sentinel Online—Sentinel Online. Available online: <https://sentinel.esa.int/en/web/sentinel/missions/sentinel-2> (accessed on 27 June 2022).
56. NOAA Shoreline Website. Available online: <https://shoreline.noaa.gov/data/datasheets/cusp.html> (accessed on 29 June 2022).
57. Aslaksen, M.L.; Blackford, T.; Callahan, D.; Clark, B.; Doyle, T.; Engelhardt, W.; Espey, M.; Gillens, D.; Goodell, S.; Graham, D.; et al. Scope of Work for Shoreline Mapping under the NOAA Coastal Mapping Program, Version 15. Available online: <https://geodesy.noaa.gov/ContractingOpportunities/cmp-sow-v15.pdf> (accessed on 13 April 2023).
58. PEPS—Operating Platform Sentinel Products (CNES). Available online: <https://peps.cnes.fr/rocket/#/home> (accessed on 29 June 2022).
59. Sen2Cor—STEP. Available online: <http://step.esa.int/main/snap-supported-plugins/sen2cor/> (accessed on 31 January 2023).
60. Level-2A Algorithm—Sentinel-2 MSI Technical Guide—Sentinel Online. Available online: <https://copernicus.eu/technical-guides/sentinel-2-msi/level-2a/algorithm-overview> (accessed on 17 April 2023).
61. Sentinel-2—Data Products—Sentinel Handbook—Sentinel Online. Available online: <https://sentinel.esa.int/en/web/sentinel/missions/sentinel-2/data-products> (accessed on 31 January 2023).
62. Cavone, G.; Giaquinto, N.; Fabbiano, L.; Vacca, G. Design of Single Sampling Plans by Closed-Form Equations. In Proceedings of the 2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Minneapolis, MN, USA, 6–9 May 2013; pp. 597–602.
63. Nicholson, W.L. On the Normal Approximation to the Hypergeometric Distribution. *Ann. Math. Stat.* **1956**, *27*, 471–483. [[CrossRef](#)]
64. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.