

Article

Detection and Monitoring of Volcanic Islands in Tonga from Sentinel-2 Data

Riccardo Percacci , Felice Andrea Pellegrino  and Carla Braitenberg * 

Department of Mathematics, Informatics and Geosciences, University of Trieste, Via Weiss, 2, 34128 Trieste, Italy; riccardoguidopowell.percacci@phd.units.it (R.P.); fapellegrino@units.it (F.A.P.)

* Correspondence: berg@units.it

Highlights

What are the main findings?

- Development of a semi-automated procedure for the detection of changes in volcanic island sub-aerial extension, leveraging Sentinel-2 data using the Google Earth Engine platform.
- Application of the method focused on three volcanic islands using multi-year data, and on a large region (Tonga Arc) with parallelized code deployed on an HPC cluster.

What is the implication of the main finding?

- The method successfully detected important changes caused by volcanic events at selected sites. Computational feasibility is assessed.
- The tool offers potential applications for navigation safety and environmental monitoring.

Abstract

This work presents an automated method for detecting and monitoring volcanic islands in the Tonga archipelago using Sentinel-2 satellite imagery. The method is able to detect newly created islands, as well as an increase in island size, a possible precursor to an explosion due to magma chamber inflation. At its core, the method combines a U-Net-type convolutional neural network (CNN) for semantic segmentation with a custom change detection algorithm, enabling the identification of land–water boundaries and the tracking of volcanic island dynamics. The algorithm analyzes morphological changes through image comparison and Intersection over Union (IoU), capturing the emergence, disappearance, and evolution of volcanic islands. The segmentation model, trained on a custom dataset of Pacific Ocean imagery, achieved an IoU score of 97.36% on the primary test dataset and 83.54% on a subset of challenging cases involving small, recently formed volcanic islands. Generalization capability was validated using the SNOWED dataset, where the segmentation model attained an IoU of 81.02%. Applied to recent volcanic events, the workflow successfully detected changes in island morphology and provided time-series analyses. Practical feasibility of the methodology was assessed by testing it on a large region in Tonga, using an HPC cluster. This system offers potential applications for geophysical studies and navigation safety in volcanically active regions.



Received: 31 October 2025

Revised: 11 December 2025

Accepted: 19 December 2025

Published: 23 December 2025

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: volcanic islands; Tonga archipelago; Sentinel-2; semantic segmentation; change detection; convolutional neural network; environmental monitoring

1. Introduction

Oceanic islands are the sub-aerial summits of submarine volcanoes (seamounts) that rise from the ocean floor, typically from depths of 1000–4000 m below sea level [1]. When such volcanoes are active, dynamic phenomena such as magmatic outflow and violent explosions may occur. In these scenarios, seamounts can emerge from sea level to form visible islands, the growth and collapse of which may occur within a few months or weeks. Volcanic activity of seamounts poses significant threats, for which maritime traffic, aviation, and residents are alerted on a 4-level scale, defining the minimum distance that must be kept from the island [2].

On existing ocean islands, one premonitory sign of volcanic unrest is ground uplift, which manifests as an increase in sub-aerial land surface. Such morphological changes can be effectively observed and monitored using satellite instruments. The objective of this study is to develop an automatic method for detecting and monitoring changes on oceanic islands, a capability that could become a vital component of a global volcanic hazard early warning system. The short time resolution of satellite images (5 days for the Sentinel-2 constellation) has the potential to enable the issuing of warnings for navigation, where knowledge on the location of shoals, islands, and volcanic hazards is of primary importance.

Existing literature already provides automated deformation (InSAR) [3] and thermal anomaly [4] monitoring. However, these systems are optimized for continental volcanoes and large volcanic centers. Small oceanic islands present unique technical challenges, such as mixed land–sea pixels, tidal and sea-level effects, and rapid coastal morphology changes. To address these limitations, this work develops a near-real-time approach tailored to detecting the emergence and evolution of seamounts and volcanic islands, taking advantage of the 10 m spatial resolution of Sentinel-2 imagery in combination with land–water segmentation. The utility of Sentinel-2 for mapping shoreline changes has been demonstrated in several recent studies [5,6].

In this study, we present an automated workflow for monitoring the emergence and morphological evolution of volcanic islands in the Tonga Archipelago. The procedure analyzes Sentinel-2 multispectral imagery to extract land–water boundaries via convolutional neural network (CNN) semantic segmentation, and applies a custom change detection module to identify land changes between pairs of images. To show the potential of our methodology, we analyze multi-year data focused on Metis Shoal and Home Reef islands, belonging to the Kingdom of Tonga. We also test the procedure over a polygonal area of 24,000 km² in the Tonga Archipelago, to assess operative feasibility on a larger scale.

2. Related Work

Land–water segmentation of remote sensing multispectral images has been a topic of research for many years. Among the simplest methods are thresholding techniques based on indices such as NDWI (Normalized Difference Water Index) or NDVI (Normalized Difference Vegetation Index) [7]. The formula for NDWI is

$$\text{NDWI} = \frac{\text{Green} - \text{NIR}}{\text{Green} + \text{NIR}} \quad (1)$$

where Green and NIR represent a Green and a Near Infrared band, respectively. NDWI is a sufficient method in a number of scenarios; however, it has important limitations hindering larger-scale applications. It is overly sensitive to vegetation and is inaccurate with shadows. Moreover, optimal thresholds vary greatly between regions, as well as due to weather and light conditions [8]. NDVI is expressed similarly to NDWI, as the normalized difference between a Red and Near Infrared band. Although primarily used to detect the presence of vegetation, it can also be applied to distinguish land from water,

but suffers from similar limitations as NDWI. Learning-based algorithms provide better solutions for the segmentation of remote sensing images. Recently, deep learning methods have become a research topic in this field.

Wieland et al. [9] compared the use of U-Net [10] and DeepLabV3+ [11] with different choices of encoder backbones (MobileNet-V3, ResNet-50, and EfficientNet-B4), applied to water body segmentation of RGB images captured by satellite and airborne cameras. The study suggested U-Net with MobileNet-V3 as the best architecture, based on their test data. They observed that the addition of an Infrared band, if available, slightly improved the accuracy. Sun et al. [12] proposed a model architecture based on DeepLabV3+, with a new fusion mechanism for high- and low-level features. In their network—which is intended for segmentation of lakes and rivers—they re-designed the Atrous Spatial Pyramid Pooling (ASPP) module to improve discrimination between adjacent objects of similar colors.

In 2018, Hu et al. [13] proposed a novel residual network block architecture, named Squeeze-and-Excitation (SE) blocks. These adjust channel-wise feature responses by capturing and leveraging the relationships between different channels. The Squeeze mechanism tackles the filers' limited receptive fields by pooling global information into a channel descriptor. This descriptor is used in the Excitation step via a gating mechanism. SE blocks can be added to a residual network, helping it focus on specific regions based on channel dependencies. Zhang et al. [14] applied an SE Residual Network (SE-ResNet) to land cover segmentation of high-resolution remote sensing images, proving it more powerful than other networks such as U-Net, ResNet50, and DeepLabV3+. SE-ResNet can also be employed as an encoder backbone for U-Net.

A different attention mechanism was introduced in Oktay et al. [15]. In their new architecture, called Attention U-Net, they employ Attention Gates (AG), which learn to suppress irrelevant regions of the image. This unit is able to focus on target objects of different shapes and scales. The authors integrated AGs into the U-Net's skip connections, for finer control over information flow between encoder and decoder. The tests performed showed that Attention U-Net outperformed standard U-Net in medical image multiclass segmentation over different datasets and training sizes. More recently, Ghaznavi et al. [16] compared the performance of simple U-Net, Attention U-Net, and a U-Net with VGG16 encoder backbone, with the objective of extracting inland water bodies from RGB satellite images. For this goal, VGG16 U-Net had the highest accuracy scores, though Attention U-Net was very close.

Hybrid approaches to model architectures are possible. For example, Cui et al. [17] proposes a modified U-Net model, with a CNN-based encoder and a decoder based on the Mamba architecture [18]. CNN–transformer hybrids have also been explored [19], where a convolutional encoder is combined with transformer-based modules, enabling the model to capture both local spatial detail and long-range dependencies with semantic features.

3. Study Area and Data

3.1. Study Area

The study area comprises the ~2500 km long Tonga–Kermadec Trench, marking the subduction of the Pacific Plate beneath the Australian Plate [20,21] and forming a continuous subduction-arc-back-arc complex characterized by frequent and strong earthquakes of magnitude greater than M 7 [22–24] and numerous active volcanoes [21]. The arc extends northward from New Zealand's North Island (Figure 1). The rapid convergence of the Kermadec-Tonga and Pacific plates—reaching speeds of up to 24 cm per year [25]—drives the intense tectonic activity in the region. The volcanic activity comprises submarine explosive eruptions producing extensive pumice rafts, which can be observed from satellites [26], and Surtseyan activity. Caldera-forming silicic eruptions, especially in the Kermadec seg-

ment, have been documented [26,27], as well as basaltic eruptions [28]. The volcanically active region in the Tonga portion of the trench spans a region of more than 500 km in length [29]. We capture this region in a polygonal Region of Interest (ROI) of roughly 24,000 km².

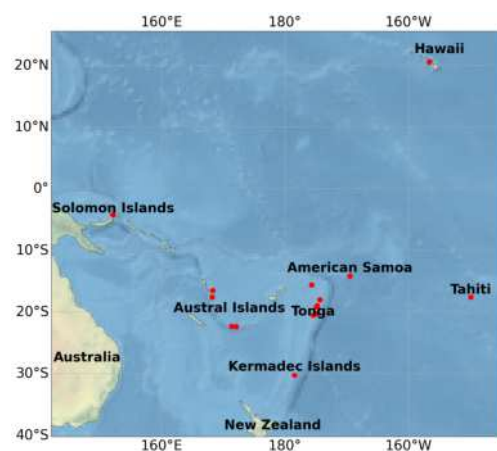


Figure 1. Distribution of dataset sample locations. Red dots indicate locations at which data were extracted.

In recent times, several volcanic islands have emerged in the Tonga archipelago, including Home Reef and Metis Shoal and others like Hunga Tonga-Hunga Ha’apai [30]. The latter increased the areal extent of the islands [6] starting from a month before the great volcanic explosion of 14 January 2022 [31,32].

Home Reef is located southwest of the Vava’u island group (Latitude -18.991° , Longitude -174.763°). It existed temporarily above sea level in 1852, 1857, and 1984, before emerging again in 2006. The island still exists and is actively evolving. In 1984, an eruption created a short-lived small island of about 0.5 km by 1.5 km. In 2006, it formed again and reached almost the same size. Several more eruptions occurred in 2022 and 2023. Recent volcanic activity came through June–July 2024 to January 2025, during which time the island virtually doubled in size. Plank et al. (2025) [33] integrated multi-sensor satellite observations to manually analyze eruption dynamics at Home Reef from September 2022 to September 2024.

Metis Shoal (Latitude -19.183° , Longitude -174.867°), also known as Late’iki, is an island close to Home Reef. Reports of eruptions and ephemeral islands associated with the volcano date back as far as 1781. The most recent island lasted from 1995 to 2019. In early November 2019, an eruption created a new and bigger island about 120 m west of where the former disappeared. By mid mid-January 2020, the new island had disappeared and is still below sea level.

3.2. Data Sources

This paragraph presents the sources of data and their use within this work. All datasets are accessed via the Google Earth Engine (GEE) [34] platform, except for the SNOWED dataset [35], which was downloaded at <https://doi.org/10.5281/Zenodo.7871636>. Sentinel-2 is a Multispectral Satellite constellation launched by the European Space Agency (ESA) in association with the Copernicus Programme in 2015, which acquires optical imagery at visible to shortwave-infrared wavelengths at high resolution (10 m to 60 m). Sentinel-2 Level-1C imagery is used to form the training dataset for the segmentation model. It is also the source of data for the automatic change detection algorithms. This Sentinel product comes with 13 spectral reflectance bands scaled by the factor 10,000.

Within the raster of Sentinel-2 products, the Cloud Score+ dataset gives cloud scores to individual pixels, and is used to gather cloudiness statistics pixel- and region-wise. Cloud scores define the usability of pixels, and are given as continuous values from 0 to 1, where 0 represents “not clear”, and 1 represents “clear” pixels. For this work, we choose to set 0.7 as the minimum value for a pixel to be considered “clear” or not occluded. The MYD14A2.061: MODIS Aqua Thermal Anomalies and Fire 8-Day Global 1 km dataset was used for the observation of volcanic activity indicators. This dataset provides 8-day fire mask composites at 1 km resolution. The fire mask features nine classes, of which three relate to Fire (heat), with low, nominal, and high confidence associated. The Thermal Infrared wavelengths used have greater cloud-penetrating ability and thus are useful in filling temporal gaps where volcanic smoke occludes Sentinel-2 images. Finally, SNOWED is a collection of 4334 Sentinel-2 Level-1C images from all continental U.S. shorelines with portions of Alaska, Hawaii, the U.S. Virgin Islands, Pacific Islands, and Puerto Rico. Each image is accompanied by a labeled version, with land/water class assigned to every pixel. By measuring performance on the SNOWED dataset, we can obtain an idea of how well each model generalizes to a broader context. All datasets are summarized in Table 1.

Table 1. Data sources used for this work.

Dataset	Description
Harmonized Sentinel-2, Level-1C	13-band Multispectral Instrument
Cloud Score+ S2_HARMONIZED V1	Pixel-level cloudiness score, based on Sentinel-2 Level-1C
MYD14A2.061	MODIS Thermal Anomalies & Fire 8-Day Global 1 km
SNOWED (Sentinel2-NOAA Water Edges Dataset)	Labeled shoreline images from the U.S. and other locations

3.3. Training Dataset Construction

In order to train the CNN for segmentation, a data set of 424 images with a window of 256×256 pixels is collected from Sentinel-2 Level-1C images [36]. Using the GEE API, these images were selected from the Tongan archipelago, as well as from other parts of the Pacific, namely Hawaii, Papua New Guinea, New Caledonia, Samoa, Tahiti, and Vanuatu (Figure 1). All selected images feature coastal scenes and form a representative sample of the typical morphology in the region of interest. Ground truth for these images was manually annotated. Multi-channel images were extracted as multidimensional numpy arrays and exported along with the respective ground truth arrays. Ground truth arrays are binary arrays where class 0 indicates water and class 1 indicates land.

A few regions are captured at more than one moment in time. The reason for this is two-fold. Firstly, the spectral response of land, as well as sea, has a different shape depending on season. We therefore choose to reuse some regions and geometric features to represent this variability. Secondly, some of the selected regions capture small volcanic islands (such as Metis Shoal or Home Reef), which represent an important target group within this work. Capturing such islands at various moments in time also relies on their often rapidly evolving morphology. Each image in this group has a copy, which is augmented via a periodic translation in both directions. All images are also randomly rotated at 0, 90, 180, and 270 degrees.

The data set is split into Training, Validation, and Test sets, using a 64%, 16%, 20% split. The split aims at providing enough training data for complex feature learning, while maintaining a sufficiently large test set to reliably assess model generalization, and is quite common in computer vision and remote sensing applications [37–39]. First, the data is divided into Training and Non-Training subsets. Then, the Non-Training subset is further split into Validation and Test sets. At each stage, elements are selected at regular intervals (every n -th element), with n chosen according to the size of the split. All Sentinel-2 bands

are included in the dataset. A discussion on the choice of bands for segmentation is included in Section 4.1.2.

From the initial annotated dataset, we also consider the set of images containing small volcanic islands of recent formation. The islands represented in this subset are Home Reef, Metis Shoal, and Hunga Tonga-Hunga Ha’apai (after the 15 January 2022 eruption). The subset, hereafter called “Small Islands” represents an important target group within this work, and is used for testing models in critical scenarios. The size of the set is 103, and it is evenly distributed between Train, Validation, and Test sets (Table 2).

The initial dataset contains some images with clouds. We opt to include them in order to train models to handle light cloudiness, which is to be expected in practical applications. These images can then be removed in the testing phase, in order to gauge the robustness of the models to cloud presence.

The data exhibits a slight imbalance between land and water pixels, with water being the predominant class. The imbalance is much stronger in the Small Islands subset ($\approx 99\%$ of water). Table 2 sums up statistics relative to the three datasets.

Table 2. Dataset statistics. Percentages indicate the fraction of Small Island and Cloudy images within each subset. Land and water fractions are averaged across images in each set.

Dataset	Size	Small Islands (%)	Cloudy (%)	Land Fraction	Water Fraction
Training	271	24.35	17.34	0.31	0.69
Validation	68	23.53	19.12	0.29	0.71
Test	85	24.71	17.65	0.32	0.68
Small Island subset	103	100	18.45	0.01	0.99
Overall	424	24.29	17.69	0.31	0.69

4. Methods

The proposed workflow accesses Sentinel-2 Level-1C imagery through the GEE API. The user-defined ROI is subdivided into 256×256 pixel tiles, which are processed iteratively. For each tile, the two most recent cloud-free images are retrieved, and land–water segmentation is applied. A dedicated change detection module then evaluates whether significant changes in subaerial land area have occurred between the two images. Figure 2 illustrates this Section’s structure.

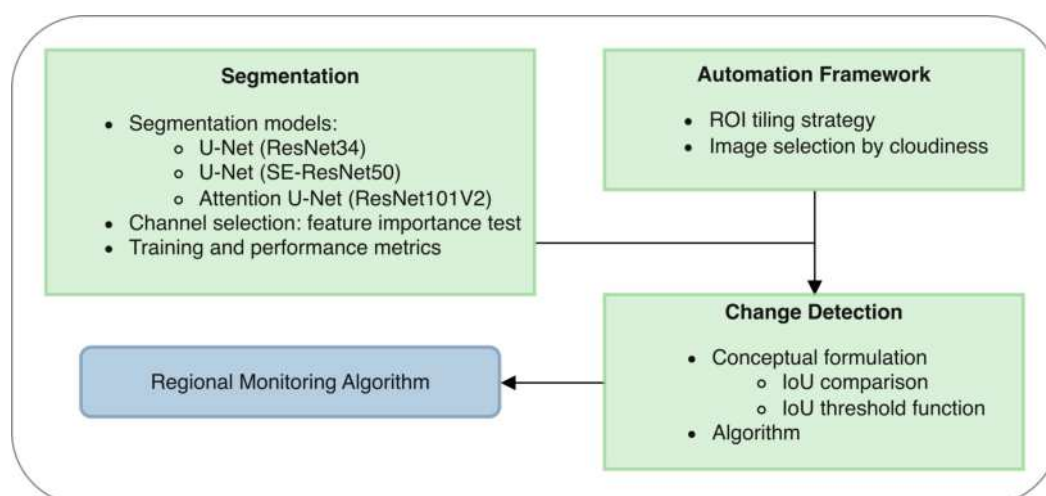


Figure 2. Schematic workflow illustrating the technical framework and processing steps, resulting in the algorithm for automated regional monitoring.

4.1. Segmentation

4.1.1. Segmentation Models

We use a U-Net-type model for segmenting multispectral images. U-Net is one of the most widely used models for semantic segmentation and is considered a benchmark among computer vision models. It is usually applied to gray-scale or RGB images, but can be extended to work with multi-channel images. U-Net is a Fully Convolutional Neural Network (FCN), based on an encoder-decoder structure. The encoder is a typical convolutional network consisting of a series of convolutions, each followed by a Rectified Linear Unit (ReLU) and a max pooling operation. Through this downsampling path, the image's size is decreased while its number of channels is increased. In other words, spatial information is reduced, and feature information is increased, helping the network learn dense image features and capture context. In the symmetric decoder block, pooling operations are replaced by up-sampling operations, reconstructing the desired resolution. Their output is concatenated with features from the corresponding encoder layer, through so-called skip connections. Skip connections are one of U-Net's defining features, and are crucial in recovering spatial details that would be lost due to downsampling. Different backbone architectures can be inserted as the encoder.

For this study, the models tested are U-Net with ResNet34 backbone, as well as U-Net with SE-ResNet50 backbone and Attention U-Net with ResNet101V2 backbone. The last two models feature attention mechanisms, as explained in Section 2. ResNetV2 [40] is a variation on ResNet [41], which changes the order of operations within the residual blocks, improving the way data flows through the network. This facilitates the training of very deep networks by allowing the gradients to flow more easily during backpropagation. In this work, the backbone is a lightweight version of the full ResNet101V2, which shares the same design principles while cutting down on model complexity.

A diagram of the Attention U-Net architecture adopted is shown in Figure 3. The structure of U-Net with SE-ResNet50 backbone is similar to that of U-Net, but the plain encoder is substituted by an SE-ResNet50 encoder, adding residual blocks and Squeeze-and-Excitation blocks. All models used have an encoder depth of 5, but the filter sizes are (256, 128, 64, 32, 16) for the ResNet34 and SE-ResNet50 variants and (1024, 512, 256, 128, 64) for Attention U-Net. The number of trainable parameters for U-Net with ResNet34 backbone, U-Net with SE-ResNet50 backbone, and Attention U-Net is respectively 24.5 million, 35.1 million, and 36.8 million.

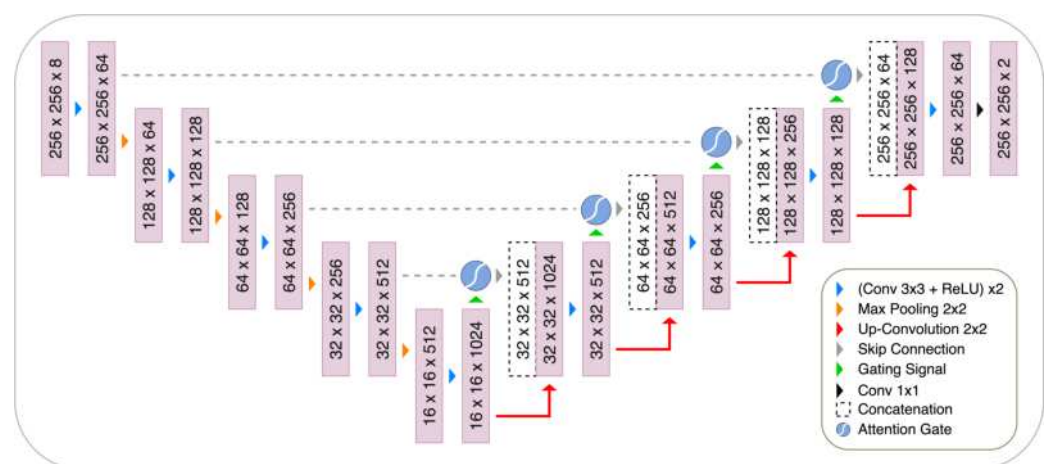


Figure 3. Diagram of the Attention U-Net architecture employed in this work. Inputs are 256×256 pixel images with 8 channels. The shape of the data going through the network is represented in the boxes as triplets of width, height, and number of features. The final layer uses softmax activation, producing a 2-channel image for binary classification. Adapted from [15].

4.1.2. Training Setup and Metrics

The models are implemented in Python using the Keras framework [42]. The U-Net implementation is taken from the `segmentation_models` library (version 1.0.1) [43], while Attention U-Net is taken from the `keras-unet-collection` library (version 0.1.13) [44]. Training is done using early stopping to prevent overfitting (`keras.callbacks.EarlyStopping`), with a patience of 40 or 50 epochs, and a warm-up stage of 20 epochs. A second callback (`keras.callbacks.ModelCheckpoint`) is used to save the best model in real time, which allows for saving more models than just the best overall. In both callbacks, the metric to monitor is validation loss. A batch size of 16 is used.

To assess the relative contribution of individual spectral bands and vegetation/water indices to model predictions, we apply a gradient-based feature importance analysis. Specifically, we compute the channel-wise mean absolute gradient of the model output with respect to each input band, a method commonly used to approximate input sensitivity in deep learning models [45,46]. To ensure comparability across bands and models, we normalize the feature importance vectors from each trained model to unit norm. We train four instances of a baseline U-Net model with a ResNet34 backbone, and compute a normalized importance vector for each. We then average these vectors to obtain a more robust estimate of per-band importance (Figure 4).

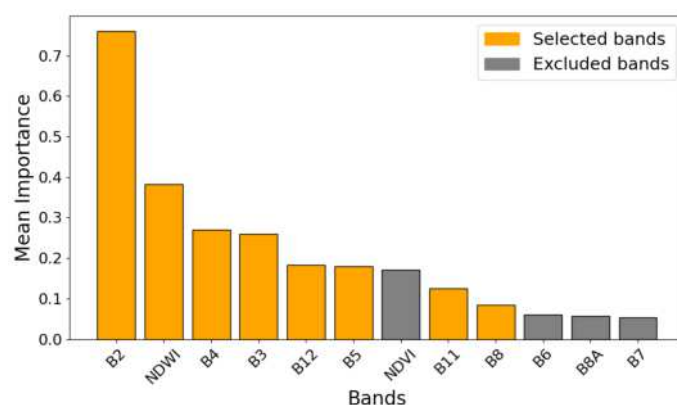


Figure 4. Mean normalized feature importance of Sentinel-2 spectral bands across four independently trained U-Net models, in descending importance order. Bars show mean importance values. Bands highlighted in orange were selected for model development.

The resulting mean feature importance vector is used to guide the selection of a compact yet spectrally informative subset of Sentinel-2 bands. Specifically, we select all bands except B6, B7, B8A, and NDVI, for a total of 8 bands. This choice is made to retain the full core spectral range of Sentinel-2, while minimizing redundancy from adjacent or low-importance channels. B6, B7, and B8A are excluded due to their lower relative importance in the gradient-based analysis, whereas B5 and B8, which fall in the same VNIR spectral region, are retained. NDVI is excluded from the selected subset due to its limited relevance to the primary task of island segmentation. In contrast, NDWI, which emphasizes land–water boundaries, is retained due to its higher feature importance and direct relevance to distinguishing coastal and aquatic regions.

A weighted Binary Cross-Entropy (BCE) loss is adopted to improve segmentation of small volcanic features such as newly emerged seamounts. These create challenges in segmentation, due to their small size, and their spectral similarity with water. Such thematic features are collected in a dataset (see Section 3) to monitor the performance of models on this important task. Weight maps are constructed such that all pixels on small islands receive a weight of 8, while in other land regions, only shoreline pixels are up-weighted to 5 and all others retain a weight of 1. Shoreline pixels are defined as

land pixels whose 11×11 square neighborhood includes at least one water pixel. The loss function is calculated by element-wise multiplication of the BCE loss matrix with the weight map matrix.

In order to evaluate the performance of the models under consideration, we adopt the following performance metrics: Precision, Recall, F1 score, Intersection-over-Union score (IoU), and Cohen's kappa. Precision is the fraction of relevant instances among the retrieved instances, while Recall is the fraction of relevant instances that have been retrieved. Using True Positives (TPs), False Positives (FPs), and False Negatives (FNs), these are formulated as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

The F1 score is the harmonic mean of Precision and Recall:

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

This provides a balanced view of segmentation accuracy, which is useful in the case of imbalanced class distributions such as in our data.

IoU (or Jaccard index) is used in different fields to measure the similarity of sample sets. In our discussion, IoU is scaled by 100, and is defined for two sets A and B as follows:

$$\text{IoU}(\%) = \left(\frac{|A \cap B|}{|A \cup B|} \right) \times 100, \quad (5)$$

where $|S|$ denotes the cardinality of set S . In our case, IoU is associated with the land class, and therefore measures the proportion of land overlap relative to the total land area, over two images. For validating the segmentation models, we consider IoU between the predicted segmentation map and the ground truth mask. The closer this value is to 100, the more accurate the segmentation.

Finally, Cohen's kappa quantifies the agreement between predicted and reference labels while accounting for chance agreement. It ranges from -1 (complete disagreement) to 1 (perfect agreement), with 0 indicating agreement equivalent to random chance. Its formulation is

$$\kappa = \frac{p_0 - p_e}{1 - p_e}, \quad (6)$$

where p_0 is the observed agreement, i.e., the fraction of pixels where the predicted and reference classes match, and p_e is the expected agreement by chance, computed from the marginal distributions of the two classifications.

All metrics, except for Cohen's kappa, are calculated relative to the land class. As shown in Section 3.3, land pixels represent a smaller fraction of each image than water. Measuring performance over the water class—for example, using Overall Accuracy or mean IoU—would inflate scores and mask errors in detecting land. Focusing on the land class, therefore, provides a more meaningful assessment of the model's ability to accurately delineate island regions, which is the primary objective of the segmentation task.

4.2. Automation Framework

4.2.1. Tiling

The region of interest is tessellated into 256×256 tiles, in order to be given to the trained segmentation model. Optionally, the set of tiles can be duplicated at a distance of 128 in both directions and both orientations, to include overlapping and improve the

accuracy of segmentation (areas near the edges of the image are more susceptible to errors, as part of the context is missing). We thus define a “primary” and “secondary” set of tiles. The resulting set of tiles is then used to analyze the region of interest.

The rule applied for gathering overlapping segmentation maps consists of updating the full segmentation map via a logical OR operation. In this way, each pixel is 0, unless any of the segmentation maps has a value 1 for that pixel. This choice is aimed at mitigating border effects in which land area is often underestimated when occurring across adjacent tiles.

4.2.2. Image Selection and Cloudiness Filter

Cloudy pixels are identified using the Cloud Score+ Sentinel-2 product (see Section 3.2). Clear images are selected by taking the percentage of cloudy pixels within a region of interest, and comparing it to a defined threshold. We therefore use a cloud cover threshold to filter out cloudy images. In our methodology, we apply the filter to each tile while traversing the region, thereby selecting for each tile the most recent clear image.

4.2.3. Preprocessing

Co-registration of images for this work is not feasible due to the lack of reference features in open ocean settings. We rely on the precision in the position of the published Sentinel-2 products. Preprocessing of the images for use in the change detection workflow requires resampling of the bands with 10 m grid cells. This is done with a bilinear interpolation of the four nearest pixels. In the GEE Python API, this logic is encapsulated in the script `ee.data.computePixels`.

4.3. Change Detection

4.3.1. Conceptual Formulation

To detect change between two images at different times, we consider the two corresponding segmentation maps, which are combined to produce a change mask. We should keep in mind that when more land is present (in particular with longer shorelines), more absolute change is inherent, mostly due to tidal effects. On the other hand, less absolute change is expected when dealing with very small land cover. It seems logical, then, to count the number of changed pixels and perform some normalization step. Two options are:

- Normalize the number of changed pixels relative to the size of the landmass. Consistency and sensitivity to varying land area.
- Calculate IoU for class land between the two segmentation maps and set a threshold for evaluation. IoU represents the proportion of land overlap relative to the total area classified as land: a measure of landmass similarity between the images.

We choose to use IoU thresholding as the change detection method. It is preferred because of its inherent normalization and compactness. Accordingly, we define the change detection criterion between two segmentation maps A and B as follows:

$$\text{IoU}(A, B) \leq T,$$

where T is a threshold value.

By looking at IoU results from the segmentation of sample images, we observe how smaller islands often have relatively low IoU values even when maintaining the same real shape from one image to the next. These cases can be attributed to model behaviour, which is sensitive to contextual features such as brightness, color, and boundary information, e.g., amount of ocean break (waves breaking near and crashing on the shore). In several cases, the segmentation maps for images of Metis Shoal shifted the island by a short length

or reduced/increased its scale slightly. Given that consistency is less at lower island sizes, we opt to model the IoU threshold as an increasing function of land area per single tile.

The threshold is defined by the parametric logarithmic function

$$T(x, A, B) = A \log(Bx + 1),$$

where x represents the land area in pixels. With this choice of function, the threshold when no land is present is 0, in which case any appearance of land will give a positive change classification (since both IoU and threshold will be zero). In cases where land is present in the earliest of the two images, the threshold value depends on the parameters A and B .

These are set by fitting $T(x, A, B)$ to a synthetic dataset, consisting of 700 automatically generated pairs of island masks, annotated with land area x , IoU, and a manually assigned change label indicating the presence or absence of meaningful landmass change. Synthetic islands are generated by growing land pixel by pixel from a random seed, adding pixels adjacent to existing land to form contiguous regions. After a fixed number of additions, enclosed water areas are filled to complete the landmass. Changes are introduced via morphological operations (dilation, erosion), addition of new land, or removal of existing land. Labels are assigned through visual inspection of each pair. Sixty-three data points related to real cases of volcanic islands were added as well. The threshold model was trained using a cross-entropy loss with a sharp sigmoid applied to $T(x) - \text{IoU}$, treating cases where $\text{IoU} \leq T(x)$ as indicative of change. The sigmoid sharpness parameter (α) was tuned by maximizing the F1 score over the training data, resulting in an optimal value of $\alpha = 5$ (Figure 5).

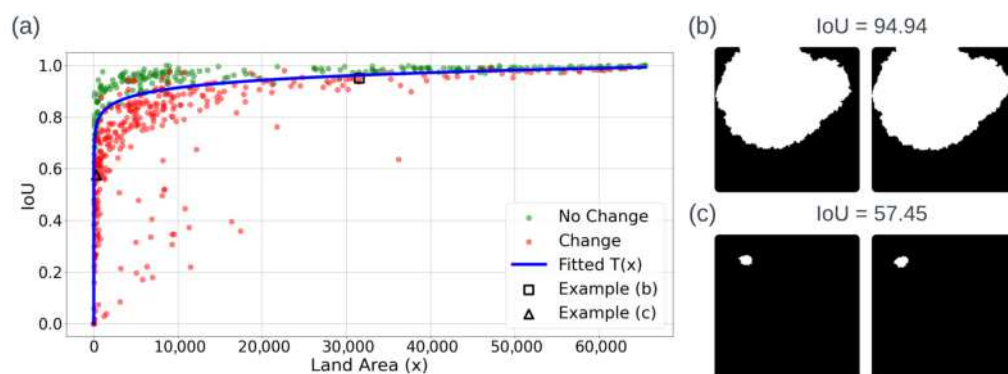


Figure 5. (a): Threshold function fitted to the synthetic dataset. The scatter plot shows elements in the synthetic dataset as Land Area-IoU pairs. Points labeled with positive (negative) change class are colored in red (green). The blue plot represents the logarithmic function T , fitted to the data. (b,c): Before and after image pairs (left to right) from the synthetic islands dataset. (b): A small mass is added to the right of the original island. While IoU is high, the new mass represents a significant change. (c) The island size is similar, but the shape and position slightly change. While we try to allow less consistency with small islands, the change is visually significant. Thus, both pairs are labeled with positive change class and are highlighted in the plot with a square (b) and a triangle (c).

4.3.2. Algorithmic Implementation

The main component of the methodology involves comparing two images taken at different times to determine if a change has occurred. The size of the images may vary if overlaps are adopted. Figure 6 shows the flowchart for the change detection algorithm.

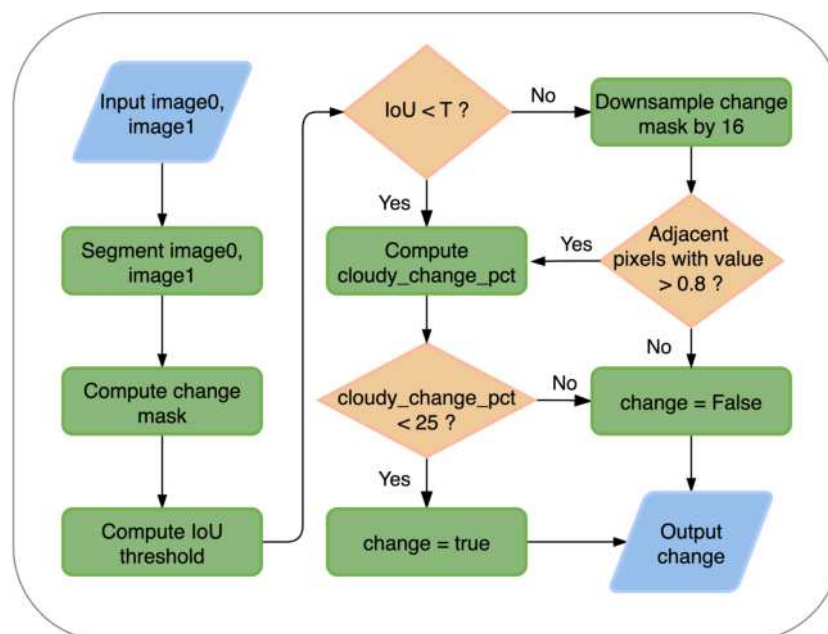


Figure 6. Change detection flowchart. IoU: IoU between image0 and image1, T: IoU threshold as a function of land size in image0, cloudy_change_pct: percentage of changed pixels which are cloudy.

The boolean output variable is named change. The algorithm segments both images using a trained segmentation model and calculates the IoU for class 1 (land) from the segmentation maps. The threshold for IoU is calculated via the logarithmic function defined above. The threshold is modeled as a function of the total land area in the first image, assuming the size of the image is exactly 256×256 . To apply this function to images of different sizes (if overlaps are adopted), we calculate the average land area per 256×256 tile by dividing the total land area by the coverage in terms of single tiles, i.e.,

$$\bar{A} = \frac{A_{\text{tot}}}{P} \times 256^2, \quad (7)$$

where \bar{A} is the average land area per tile, A_{tot} is the total number of land pixels (across all tiles), and P is the total number of unique pixels covered by all tiles. We can now compare IoU with the threshold:

(a) if IoU is smaller than the threshold, a possible change is detected. To verify if this is due to cloud presence, we calculate the percentage of pixels with detected change that are cloudy in either image, using the CloudScore+ dataset. If this percentage exceeds 25%, the detected change is taken as dubious, i.e., a possible false positive due to cloud artifacts, and we conservatively set change to False. If the percentage is below 25%, we accept the IoU as indicating a change.

(b) If IoU is greater than the threshold, we initially set change to False. Optionally, an additional step can be performed to detect small but significant morphological changes. For instance, we want to be able to reject normal tidal effects, but detect the emergence of a small land mass next to an already existing one. Both these events could produce a high IoU value, and be rejected by the steps described above.

To perform this check, we create a change mask from the segmentation maps, where the mask has a value of 0 where the maps are identical and 1 otherwise. This mask is then downsampled by a factor of 16, resulting in an image with pixel values in the range [0, 1] and pixel size 160 m (from the original 10 m). We search for pairs of adjacent pixels with high values (greater than 0.8), which indicate significant changes over areas of $51,200 \text{ m}^2$

(0.0512 km²). If such changes are identified, we perform the cloudiness verification step again to rule out cloud artifacts. If this test is passed, change is set to True.

4.4. Regional Monitoring Algorithm

We can now describe the main algorithm for monitoring a user-defined region for changes in real-time. The procedure can be executed at regular time intervals, as new images become available. The set of primary tiles covering the ROI is looped through, iteratively applying the change detection module described in Section 4.3, which fetches the two latest cloud-free images and sets a positive or negative value for change to each visited tile. Results are written to a log file, in the minimal form of the following:

- tile number: left to right, bottom to top,
- coordinates: top-left coordinates for this tile,
- last analyzed image: date of the last image that was analyzed during a run, for this tile,
- change: whether change was detected in this tile during the last run.

Other data from the run can be saved, for the purpose of analyzing the results.

We assume in the following that tiling includes overlapping “secondary” tiles. The algorithm iterates over the set of primary tiles. For each tile, we fetch a collection of images, filtering by a chosen cloudy pixel percentage threshold, and obtain the two latest images. By reading the date of the last analyzed image in the log file, we verify whether newer images are available. If none are, the tile is either skipped, or the cloudiness filter threshold is raised until a new image is found. If a new image is available, information from all tiles intersecting the central tile is included, meaning that the central tile is joined with all intersecting tiles in the secondary set, using the same Sentinel-2 image. Then, the change detection procedure described above is applied, producing a True/False change value. Results for this tile are written to the log file, and we move to the next primary tile, until all have been visited.

4.5. Parallelization

To parallelize the code, we adopted a CPU-based strategy using Python’s multiprocessing module. This approach is well-suited to our workflow, since the data can be subdivided into independent tiles that are processed without interdependence, making task parallelism on multi-core CPUs efficient and straightforward to implement. This also provides a scalable solution to be deployed on a high-performance computing (HPC) cluster, for rapid testing of the procedure. GPU acceleration should be sought for further optimization.

The procedures were run either locally (when focused on a single tile) or on the Demetra HPC cluster of the University of Trieste’s Department of Mathematics, Informatics and Geosciences. Analysis was run using a DELL PowerEdge R7525 server equipped with two AMD EPYC 7542 processors (32 cores each), 768 GB of RAM, 5.6 TB local storage, and two NVIDIA A100 GPUs. In our case, only CPU resources were used for analysis.

5. Results

5.1. Comparison of Segmentation Models

The two best models obtained are U-Net with SE-ResNet50 backbone and Attention U-Net with ResNet101V2 backbone. These two are compared here along with the U-Net with ResNet34 backbone, which can be considered a baseline model. Figure 7 shows graphs relative to the training runs for U-Net with SE-ResNet50 and Attention U-Net.

Table 3 displays performance metrics for the three models on Training, Validation, Test, and Small Islands datasets. In all cases, Attention U-Net performs better according to the IoU score. We are particularly interested in the performance over the Small Islands set,

which contains typical scenarios with small volcanic islands of recent formation. For this dataset, IoU scores are lower by 14–21 points with respect to the Test set. Looking at the improvement in IoU score of Attention U-Net over U-Net with SE-ResNet50, we obtain an improvement of 0.18 points on the Test set, and 1.56 on the Small Islands set. Thus, both models have similar accuracy rates in a general scenario, but Attention U-Net has a significant margin over U-Net for the Small Islands set. If we compare U-Net with ResNet34 and SE-ResNet50 backbones, in the Test set, IoU is lower only by 0.34 with ResNet34, whereas in the Small Islands set, performance is significantly lower with ResNet34 (6.06 percentage points difference in IoU). Thus, it appears that U-Net benefited from SE-ResNet50’s greater complexity and attention mechanism when dealing with the critical Small Islands subset.

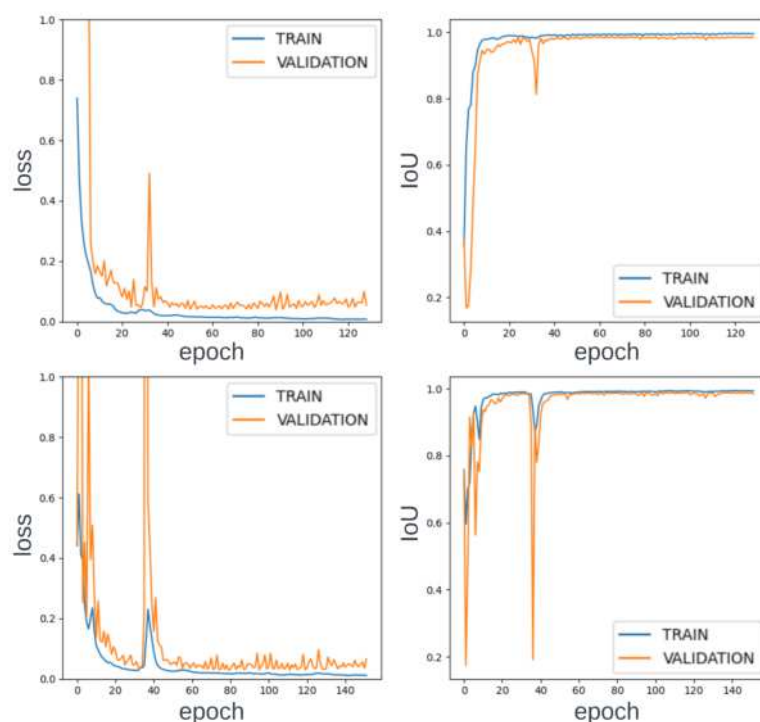


Figure 7. Training graphs for the two best models, displaying the weighted BCE loss and IoU score (here in the natural 0–1 range) for the Train set (blue) and Validation set (orange). **Top:** U-Net with SE-ResNet50, **bottom:** Attention U-Net with ResNet101V2.

Table 3. Performance metrics for the segmentation models, over the different datasets. All metrics except Cohen’s kappa are calculated relative to the land class. “CF” indicates a cloud-free dataset.

Dataset	U-Net (ResNet34)					U-Net (SE-ResNet50)					Att. U-Net (ResNet101V2)				
	Prec.	Rec.	F1	IoU	κ	Prec.	Rec.	F1	IoU	κ	Prec.	Rec.	F1	IoU	κ
Train	98.95	98.59	98.77	97.56	0.982	98.58	99.60	99.09	98.19	0.987	98.67	99.61	99.14	98.30	0.988
Validation	99.04	98.40	98.72	97.47	0.982	98.48	99.22	98.85	97.72	0.984	98.64	99.23	98.93	97.89	0.985
Test	98.46	98.33	98.40	96.84	0.976	98.39	98.75	98.57	97.18	0.979	98.58	98.74	98.66	97.36	0.980
Test–CF	99.14	98.69	98.92	97.86	0.984	98.70	99.53	99.11	98.24	0.987	98.91	99.41	99.16	98.33	0.988
Small Islands	84.61	88.09	86.31	75.92	0.862	87.18	93.21	90.10	81.98	0.900	89.86	92.23	91.03	83.54	0.910
Small Islands–CF	91.20	88.67	89.92	81.68	0.898	87.52	93.91	90.60	82.82	0.905	91.11	92.45	91.78	84.80	0.917
SNOWED	89.47	90.09	89.78	81.45	0.847	91.66	89.24	90.43	82.53	0.858	87.32	91.73	89.47	80.94	0.840

We also defined further datasets for testing, by filtering out images with clouds from the Test and Small Islands datasets. These are indicated in Table 3 as “Test–CF” and “Small Islands–CF”. Fifteen images were thus removed from the Test set, and twenty from the

Small Islands set. As should be expected, performance is improved relative to the original sets. U-Net with SE-ResNet50 gained 1.06 and 0.84 percentage points in IoU score in Test and Small Islands sets, respectively, while Attention U-Net gained 0.97 and 1.26 points (compare rows 3 and 5 to rows 4 and 6 in Table 3). Since these improvements are relatively small, we can deduce that both models are fairly robust to the presence of clouds, and that their performance does not heavily degrade in operational settings where some cloudiness is expected. On the other hand, U-Net with ResNet34 gained 5.76 points in IoU in the Small Islands set, indicating a less robust model.

Lastly, in Table 3 we also evaluate the models on the SNOWED dataset, in order to obtain an idea of how well these generalize. The three models show decent results overall on this dataset, with U-Net with SE-ResNet50 achieving the highest IoU score of 82.53. Attention U-Net performed slightly worse than U-Net with ResNet34 backbone. While Recall for Attention U-Net was highest overall, Precision was the lowest. Considering that the training data for the three models are taken from an entirely different geographical context than the SNOWED data, these models show some generalization capabilities. While SNOWED contains glaciers and fjords from the northern end of America and Alaska, our dataset includes very different features such as coral reefs, atolls, and volcanic formations.

Across all datasets, F1 score and Cohen's kappa follow the same trends as the other metrics, with consistently high values for Train, Validation, and Test sets. As expected, performance drops for the more challenging Small Islands and SNOWED datasets.

5.2. Application to Test Cases

We test the change detection mechanism by using Metis Shoal and Home Reef islands as examples. The model used henceforth is Attention U-Net. Both the islands fit in a single 256×256 tile. All Sentinel-2 data related to their existence is used to create multi-year time series for land surface and change detection. Thermal anomaly data is included for a more complete view. We report the algorithm's Precision and Recall with respect to its ability to detect true changes, for both cases within the selected time frames.

The first series of events concerns the island of Metis Shoal (Late'iki). The volcano erupted in October 2019, causing the disappearance of the former island and the creation of a new one, as a result of volcanic material being ejected and accumulating above the waterline. The new island gradually eroded and was submerged by February 2020 [47] (Figures 8 and 9b). We filter the Sentinel-2 Level-1C images, with a cloud cover threshold of 5% within the tile. We then consider consecutive pairs of images from the filtered collection, with their respective segmentation maps. For each pair, an evaluation on whether a change has occurred is given.

Figure 8 shows the land surface time series for Metis Shoal. Correct and incorrect change detections are marked in green and blue, respectively. This island had only minimal changes to its shape between 2016 and 2019 (Sentinel-2 was launched in 2015), which suggests that the island formed in 1995 might have lasted for more than 20 years [48]. Incorrect change events are detected multiple times between May 2017 and May 2019. Although the island remains unchanged, the surface area plot is unstable, due to the combined effects of Metis Shoal's very small size and the variability in wave shoaling strength. In October 2019, an eruption caused the disappearance of the original island and the emergence of a new one, before the disappearance of all land above sea level. These events are recorded correctly in the right-most part of the series. A thermal anomaly shortly before the latter event is detected, as indicated by the blue vertical lines, confirming the presence of volcanic activity. With regard to the detection of real changes in sub-aerial surface, the calculated Precision and Recall for Metis Shoal within the 2016–2020 time frame are 20% and 80% respectively. A single false negative is present at the very end of the time

series, corresponding to the first image in which Metis Shoal is completely submerged. This is due to the previous image's segmentation showing no land detected, while the island is just barely above sea level.

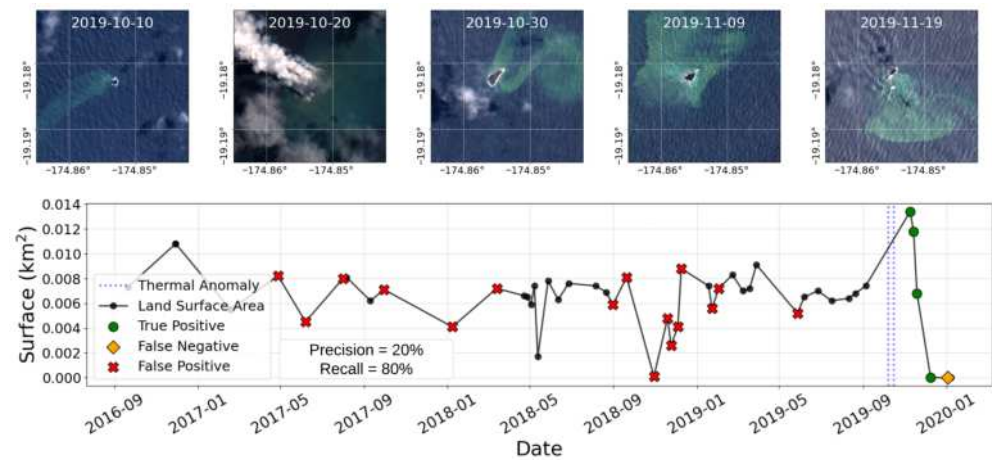


Figure 8. Top: evolution and the disappearance of Metis Shoal between October 2019 and January 2020. Bottom: Metis Shoal Surface area time series with cloudiness threshold of 5%. Green circular markers denote correct change detections (true positives), while red cross-shaped markers denote incorrect change detections (false positives). Vertical dashed lines in blue indicate the presence of thermal anomalies. Precision and Recall refer to the accuracy of the change detection. False positives in the central part are caused by segmentation inconsistency at very small island sizes. The submersion of the island starting in November 2019 is well recorded.

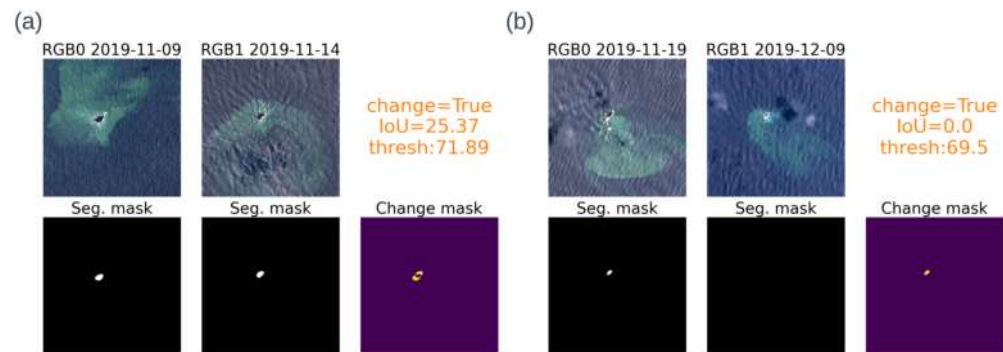


Figure 9. Change detection of two image pairs from the evolution and submersion of Metis Shoal. These show segmentation maps, change mask, change value, IoU, and threshold. Water discoloration does not interfere with segmentation.

In Figure 9, the procedure correctly detects a change between two pairs of images, as Metis Shoal evolves and submerges.

As a second test case, we consider eruptive events at Home Reef volcano. Home Reef Island underwent several changes from its emergence in September 2022. The island started changing shape in April–May 2023, until in October 2023 a series of eruptive events significantly altered its shape from a smaller, round-shaped landmass to a larger, North-South elongated one [49]. A second period of volcanic activity came recently in the summer of 2024. Weekly reports from the Smithsonian Institution state that an intense thermal anomaly was detected from satellite images on 15 June 2024, and that eruptions were ongoing between 18 June and 9 July 2024. Lava flowed from a vent on the island, expanding the land to the east [50] (Figure 10). New thermal anomalies were spotted from 4 December 2024, indicating the presence of volcanic activity which continued through spring 2025 [51]. During this period, the island continued to grow to the East, reaching virtually double its size before Summer 2024. Figure 11b shows segmentation maps for two

images from January 2025. Though segmentation is inaccurate due to smoke and cloud presence, we use an image from 21 January 2025 and focus on the island area to measure its size, obtaining a surface of 1659 pixels or 0.1659 km².

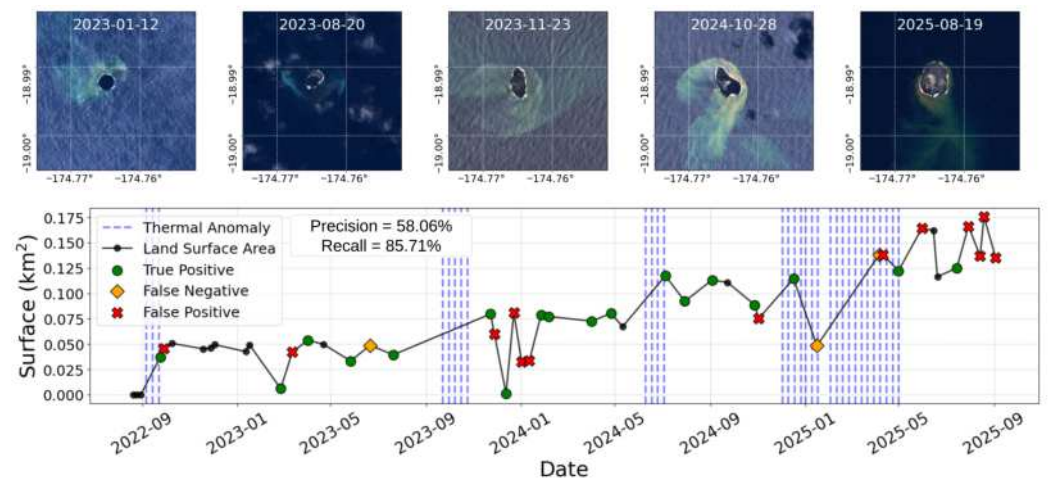


Figure 10. Top: stages of Home Reef's expansion. **Bottom:** Home Reef time series, with a cloud cover threshold of 5%. Land area is calculated by focusing on a rectangle fitted around the current shape of the island. The island shows several increments in size since its emergence. Volcanic activity leading to these is clearly marked by thermal anomalies. Some issues with clouds and smoke are present in November 2023–January 2024, as well as after Summer 2024. Data points in December 2024 through September 2025 are imprecise due to smoke and ongoing activity.

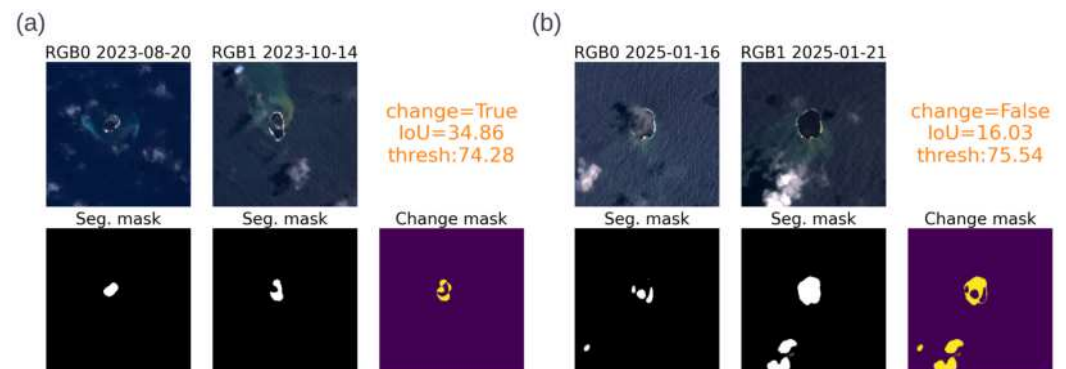


Figure 11. (a): Change detection of two image pairs with Home Reef. The cloud cover threshold here is 15%, giving more data at the cost of some noise. **(a):** The island almost doubled in size within two months. Smoke plumes and problematic volcanic surface hinder segmentation on 14 October 2023. **(b):** Recent images show the island expanded to the east. Despite obstructions from smoke and clouds, the image on 21 January 2025 yields a reliable island size estimate of 0.1659 km²—approximately double the pre-summer 2024 area.

Figure 10 is the land surface series for Home Reef between July 2022 and today. Compared to Metis Shoal, Home Reef Island continued to evolve throughout the period considered. Here, the images are filtered with a maximum cloud cover of 5%. Oscillations between December 2023 and April 2024 are due to volcanic smoke. The true size of the island during that time frame is around 0.08 km². During that period, several images are affected by smoke and cloudiness, which appear to be undetected in the Cloud Score+ cloud masks, and cause issues in segmentation. Such issues are also evident in the right-most part of the graph. Thermal anomalies occurred throughout the island's growth, testifying to intense volcanic activity that produced significant morphological changes. Precision and Recall for the data relative to Home Reef are respectively 58.06% and 85.71%.

5.3. Application to Large Region in Tonga

We apply the procedure on a large region covered by $3655 \times 256 \times 256$ pixel tiles, or $23,953 \text{ km}^2$. The region spans the Tofua Ridge, which is parallel to and some 150 km west of the Kermadec–Tonga trench (Figure 12). The Tofua Ridge represents the most active area in the Tonga section of the Kermadec–Tonga trench, and includes the volcanoes of Metis Shoal, Home Reef, Hunga Tonga–Hunga Ha’apai, Tofua, and Fonuafo’ou [52]. The area consists primarily of open ocean, but for a few islands. The test is run in order to measure the runtime of the code, and observe the behavior of the algorithm in open ocean settings.

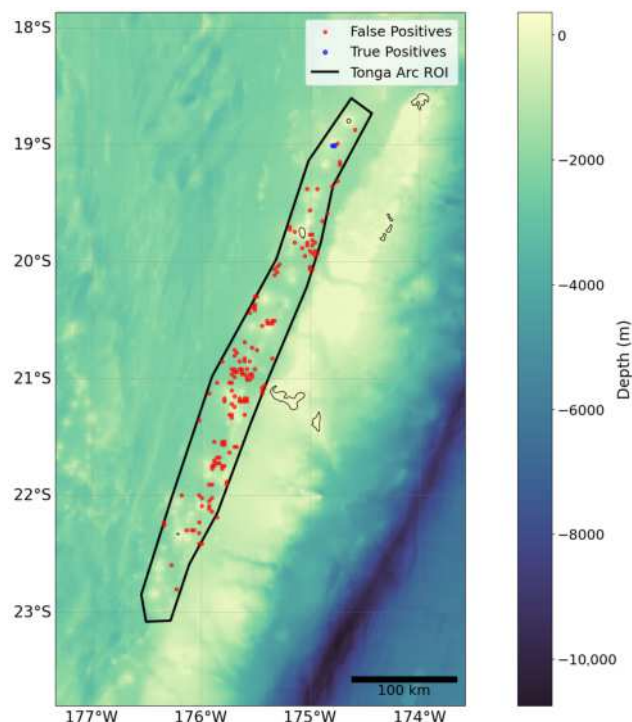


Figure 12. Area used for testing the automatic procedure on a large scale. Red dots indicate tiles that produced false positive change detections. These were mostly due to cloud presence. The two blue dots correspond to Home Reef tiles, where the change was correctly detected.

CPU-based parallelism described in Section 4.5 was used, with 24 processes spawned. The date of 24 November 2023 was used to capture an expansion in the island of Home Reef. We run the procedure twice, with and without a set of secondary overlapping tiles. The addition of the latter increases the number of CNN inferences, but mitigates edge-effects.

The number of tiles for which data was unavailable, due either to missing Sentinel-2 coverage or to cloudiness, is 73. The number of tiles which were analyzed during the procedure is thus 3582. Total runtimes with and without overlapping tiles were respectively 1 h 44 min 5.2 s (1.74 s per tile) and 45 min 14.45 s (0.76 s per tile). The observed runtimes indicate that the procedures are feasible for routine application in a cost-effective manner.

Five hundred twenty-seven analyzed tiles are considered bad data, since the relative images lie at the boundary of the Sentinel-2 swath. At these margins, missing data regions appear as black areas in the imagery and lead to segmentation issues. Among the 3055 analyzed tiles with good data, the False Positive Rate is 5.01%. Such cases are mostly due to pixels with undetected cloudiness being classified as land. The change event at Home Reef island was successfully detected.

6. Discussion

This work proposes an automatic procedure for the detection of new volcanic islands in the Tonga archipelago region, and the monitoring of their surface through time. At the core of the work is a U-Net type convolutional neural network for semantic segmentation of 256×256 pixel Sentinel-2 images. While simple methods for land–water segmentation—such as NDWI thresholding—exist, these are often inadequate. Convolutional neural networks represent a more principled choice of architecture for this problem, as they are able to leverage contextual information contained in each pixel’s surrounding.

The models explored are variants of the famous U-Net architecture, introduced in Ronneberger et al. [10]. The two best models obtained are U-Net with SE-ResNet50 backbone and Attention U-Net with ResNet101V2 backbone. Both architectures include attention mechanisms, which direct the model to focus on relevant image features. In SE-ResNet50, Squeeze-and-Excitation (SE) blocks calibrate channel-wise feature responses by explicitly modelling interdependencies between channels. In the case of Attention U-Net, Attention Gates (AG) are inserted in skip connections, and learn to suppress irrelevant portions of the image. Both models have similar complexity in terms of the number of trainable parameters.

The models were trained on a dataset that we collected using the GEE Python API. The dataset contains numpy arrays of 424 12-channel Sentinel-2 Level 1C images, with their respective ground truth arrays and weight map arrays. Ground truth arrays were manually annotated by identifying land and water regions within the images. Weight maps were applied in the calculation of a weighted Binary Cross-Entropy loss, which targets the issues specific to the set task and geography. The models were trained using a selection of the available channels, with the addition of NDWI, based on a feature importance test.

The trained models were compared on the Test set and on a subset of the whole dataset consisting of small volcanic island images. The IoU accuracy scores were good overall, but we preferred Attention U-Net for its overall edge in performance, especially on the Small Islands set. Less consistency was observed when dealing with very small islands such as Metis Shoal. Barrier reefs, typical of the Tonga region, sometimes caused nearby pixels to be classified as land, most likely due to the presence of an ocean break. These artifacts are functionally reasonable as they are associated with features typical of the land–water interface.

Change detection between two images was defined using the segmentation masks. Our proposed approach involves calculating the IoU of the two segmentation maps and determining if it falls below a certain threshold. The threshold is an increasing logarithmic function of land size, which allows more inconsistency in smaller islands. When IoU is lower than the threshold, the method verifies whether the change detection may be due to cloud presence. Change detection can also be done by downsampling the change mask to look for regions with a high density of changed pixels. The change detection method is applied to the monitoring of a region of interest of variable size. A tiling process is needed to obtain images of size 256×256 . Once the tiling is defined, the region is analyzed by looping through the individual tiles and performing change detection on each one individually.

We applied the described procedure to known test cases of volcanic activity that caused changes in the shape of Metis Shoal and Home Reef islands. Although the methods were effective in identifying actual changes, their temporal resolution was at times limited by the availability of clear images. The use of MODIS thermal anomaly data can add context where data is missing, as the thermal infrared wavelengths used can penetrate some types of clouds. When using a higher cloudy pixel percentage threshold for image selection, we can obtain more data points at the cost of some noise. However, the verification step described in Section 4.3.2 often correctly detects cloud artifacts, preventing the detection of

false changes. We analyzed time series for the two islands. In all cases, known events were successfully detected. However, some noise was present due to cloudiness and smoke from volcanic activity. Although the Precision of the change detection method was low (20.00% and 58.06% for the Metis Shoal and Home Reef series), a higher Recall (80.00% and 85.71%) is essential for the detection of dangerous cases; False Positives can be managed through manual follow-up assessments. While there is still a need for human verification in dubious cases (if not in all change classification cases), this task is not particularly cumbersome when the analyzed region is not extremely large. Out of 3055 256×256 pixel tiles analyzed for the Tonga arc area, 153 were False Positives. Thus, we believe that the procedure can be successfully applied to the Tonga region, albeit in a semi-automatic fashion.

We proved the practical feasibility of the algorithm by testing it over a large area using an HPC cluster. The processing times per tile were 1.74 and 0.76 with and without overlapping tiles. The False Positive Rate of 5.01% indicates good stability over open ocean areas.

Possible Improvements

The accuracy of the method is strongly related to that of the segmentation model. Our approach leveraged U-Net-type architectures, which are recognized as strong choices for semantic segmentation. However, various and more recent architecture types can be explored, such as CNN-transformer hybrids.

The current thresholding technique for change detection can be further refined. For example, a neural network could be trained with before-and-after images and their corresponding segmentation maps to detect changes. Depending on the architecture, these images can be concatenated into a single image of shape $(256, 256, 2 \cdot C + 2 \cdot 1)$, where C is the number of channels for each image, and 1 is the added dimension for each segmentation map, or they can be processed separately in a multi-input neural network. The network could be trained to output a binary value indicating whether a significant change has occurred between the two images, or, if more spatial detail is required, it could generate a full change mask.

Furthermore, the change detection procedure could be extended to analyze sequences of images rather than just pairs. By examining a series of images, it may be possible to track more gradual changes. In this paper, however, the focus was on detecting abrupt changes resulting from volcanic activity, particularly for hazard warning purposes. Thus, we limited our approach to using image pairs. Nonetheless, the techniques presented here could be adapted for contexts where the objective is to monitor slower changes, by extension to multi-temporal data.

The segmentation models are specific to the region of Oceania. In general, it is easier to achieve strong model performance when the problem is clearly defined and restricted to a specific task or domain. However, some level of generalization can still be attempted to broaden the model's applicability. The models demonstrated moderate generalization capability on the SNOWED dataset, although not sufficient to be applied on a global scale. Training a U-Net model (or any deep learning model) for land–water segmentation on a global scale is challenging because of the extreme variability in environmental features, such as coastlines, vegetation, terrain, water types, and seasonal differences across different regions.

To enhance generalization capabilities and make the approach applicable on a global scale, it would be necessary to incorporate additional datasets. A few global datasets for land–water segmentation do exist, such as those presented in Wieland et al. [53] and Li et al. [54]. However, such datasets do not always share the same band combination, and the wavelength ranges of corresponding bands are not always consistent. It is recom-

mended to use data that include at least one infrared band in addition to RGB. If broader geographic coverage or feature diversity is needed, existing datasets can be manually expanded, as demonstrated in this work. Additionally, integrating auxiliary features such as geographic coordinates may enhance model performance across diverse environments.

Finally, it is worth considering the use of alternative satellite platforms. Sentinel-2 offers freely accessible data, but is limited in its horizontal resolution and revisit time. Newer satellite missions provide higher spatial resolution and more frequent revisits, which would increase the likelihood of obtaining cloud-free images and improve the ability to track rapid morphological changes. However, the use of such platforms is often constrained by service costs, which must be carefully considered. In its current form, this study lays the groundwork for the development of a monitoring tool for volcanic islands.

7. Conclusions

In the present study, we presented a methodology for monitoring active volcanic regions, with the goal of detecting the emergence and change of volcanic islands. The methodology was applied to two important cases: Metis Shoal and Home Reef islands. In both cases, the procedure was able to successfully capture significant events.

This application is timely, due to the ongoing intense activity at sites like Home Reef volcano. During the year 2024, Home Reef doubled in size, reaching a surface of around 165.9 m² on 21 January 2025. The use of Sentinel-2 imagery, with a revisit time of 5 days, allows fast detection of volcanic unrest. This would provide valuable data for the issuing of hazard warnings for navigation safety.

Author Contributions: Conceptualization, R.P. and C.B.; methodology, R.P., F.A.P. and C.B.; software, R.P.; validation, R.P.; formal analysis, R.P.; investigation, R.P. and C.B.; data curation, R.P.; writing—original draft preparation, R.P.; writing—review and editing, R.P., C.B. and F.A.P.; visualization, R.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the NextGenerationEU program under the Italian National Recovery and Resilience Plan (PNRR), Grant Number (CUP) J92B24001150005.

Data Availability Statement: Supplementary data associated with this article can be found at <https://doi.org/10.17632/mfc95sgrbf.3>.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AG	Attention Gate
BCE	Binary Cross-Entropy
CNN	Convolutional Neural Network
FCN	Fully Convolutional Network
GEE	Google Earth Engine
IoU	Intersection Over Union
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIR	Near Infrared
ReLU	Rectified Linear Unit

References

1. Menard, H.W. *Marine Geology of the Pacific (International Series in the Earth Sciences)*; McGraw-Hill: New York, NY, USA, 1964. [[CrossRef](#)]
2. Global Volcanism Program. Report on Home Reef (Tonga) (Sennert, S., Ed.). Weekly Volcanic Activity Report, 8 January–14 January 2025. Smithsonian Institution and US Geological Survey. 2025. Available online: <https://volcano.si.edu/showreport.cfm?vvar=GVP.WVAR20250108-243080> (accessed on 22 January 2025).
3. Lazecký, M.; Spaans, K.; González, P.J.; Maghsoudi, Y.; Morishita, Y.; Albino, F.; Elliott, J.; Greenall, N.; Hatton, E.; Hooper, A.; et al. LiCSAR: An automatic InSAR tool for measuring and monitoring tectonic and volcanic activity. *Remote Sens.* **2020**, *12*, 2430. [[CrossRef](#)]
4. Coppola, D.; Laiolo, M.; Cigolini, C.; Massimetti, F.; Delle Donne, D.; Ripepe, M.; Arias, H.; Barsotti, S.; Parra, C.B.; Centeno, R.G.; et al. Thermal remote sensing for global volcano monitoring: Experiences from the MIROVA system. *Front. Earth Sci.* **2020**, *7*, 362. [[CrossRef](#)]
5. 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](#)]
6. Braitenberg, C. Monitoring the Hunga volcano (Kingdom of Tonga) starting from the unrests of 2014/2015 to the 2021/2022 explosion with satellites Sentinel 1-2 and Landsat 8-9. *Front. Earth Sci.* **2024**, *12*, 1373539. [[CrossRef](#)]
7. Ji, L.; Zhang, L.; Wylie, B. Analysis of Dynamic Thresholds for the Normalized Difference Water Index. *Photogramm. Eng. Remote Sens.* **2009**, *75*, 1307–1317. [[CrossRef](#)]
8. Liu, Y. Why NDWI threshold varies in delineating water body from multitemporal images? In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 4375–4378. [[CrossRef](#)]
9. Wieland, M.; Martinis, S.; Kiefl, R.; Gstaiger, V. Semantic segmentation of water bodies in very high-resolution satellite and aerial images. *Remote Sens. Environ.* **2023**, *287*, 113452. [[CrossRef](#)]
10. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18, pp. 234–241. [[CrossRef](#)]
11. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818. [[CrossRef](#)]
12. Sun, D.; Gao, G.; Huang, L.; Liu, Y.; Liu, D. Extraction of water bodies from high-resolution remote sensing imagery based on a deep semantic segmentation network. *Sci. Rep.* **2024**, *14*, 14604. [[CrossRef](#)]
13. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141. [[CrossRef](#)]
14. Zhang, X.; Li, L.; Di, D.; Wang, J.; Chen, G.; Jing, W.; Emam, M. SERNet: Squeeze and Excitation Residual Network for Semantic Segmentation of High-Resolution Remote Sensing Images. *Remote Sens.* **2022**, *14*, 4770. [[CrossRef](#)]
15. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv* **2018**, arXiv:1804.03999. [[CrossRef](#)]
16. Ghaznavi, A.; Saberioon, M.; Brom, J.; Itzerott, S. Comparative performance analysis of simple U-Net, residual attention U-Net, and VGG16-U-Net for inventory inland water bodies. *Appl. Comput. Geosci.* **2024**, *21*, 100150. [[CrossRef](#)]
17. Cui, M.; Li, K.; Chen, J.; Yu, W. CM-Unet: A Novel Remote Sensing Image Segmentation Method Based on Improved U-Net. *IEEE Access* **2023**, *11*, 56994–57005. [[CrossRef](#)]
18. Gu, A.; Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv* **2023**, arXiv:2312.00752. [[CrossRef](#)]
19. Kang, J.; Guan, H.; Ma, L.; Wang, L.; Xu, Z.; Li, J. WaterFormer: A coupled transformer and CNN network for waterbody detection in optical remotely-sensed imagery. *ISPRS J. Photogramm. Remote Sens.* **2023**, *206*, 222–241. [[CrossRef](#)]
20. Bevis, M.; Taylor, F.W.; Schutz, B.E.; Recy, J.; Isacks, B.L.; Helu, S.; Singh, R.; Kendrick, E.; Stowell, J.; Taylor, B.; et al. Geodetic observations of very rapid convergence and back-arc extension at the Tonga arc. *Nature* **1995**, *374*, 249–251. [[CrossRef](#)]
21. Smith, I.E.M.; Price, R.C. The Tonga–Kermadec arc and Havre–Lau back-arc system: Their role in the development of tectonic and magmatic models for the western Pacific. *J. Volcanol. Geotherm. Res.* **2006**, *156*, 315–331 [[CrossRef](#)]
22. Mitsui, Y.; Muramatsu, H.; Tanaka, Y. Slow deformation event between large intraslab earthquakes at the Tonga Trench. *Sci. Rep.* **2021**, *11*, 257. [[CrossRef](#)]
23. Tian, D.; Wei, S.S.; Wang, W.; Wang, F. Stress drops of intermediate-depth and deep earthquakes in the Tonga slab. *J. Geophys. Res. Solid Earth* **2022**, *127*, e2022JB025109. [[CrossRef](#)]
24. Hrubcová, P.; Rastjoo, G.; Vavryčuk, V. Stress variations in southern Tonga slab derived from deep-focus earthquakes. *J. Geophys. Res. Solid Earth* **2024**, *129*, e2023JB028039. [[CrossRef](#)]

25. Zhu, Y.; Ji, Y.; Zhu, W.; Qu, R.; Xie, C.; Zeng, D. Subduction hydrothermal regime and seismotectonic variation along Kermadec–Tonga megathrusts. *J. Asian Earth Sci.* **2023**, *243*, 105532. [[CrossRef](#)]
26. Carey, R.; Soule, S.A.; Manga, M.; White, J.D.L.; McPhie, J.; Wysoczanski, R.; Jutzeler, M.; Tani, K.; Yoerger, D.; Fornari, D.; et al. The largest deep-ocean silicic volcanic eruption of the past century. *Sci. Adv.* **2018**, *4*, e1701121. [[CrossRef](#)]
27. Wright, I.C.; Gamble, J.A.; Shane, P.A. Submarine silicic volcanism of the Healy caldera, southern Kermadec arc (SW Pacific): I–volcanology and eruption mechanisms. *Bull. Volcanol.* **2003**, *65*, 15–29. [[CrossRef](#)]
28. Wright, I.C.; Gamble, J.A. Southern Kermadec submarine caldera arc volcanoes (SW Pacific): Caldera formation by effusive and pyroclastic eruption. *Mar. Geol.* **1999**, *161*, 207–227. [[CrossRef](#)]
29. Wright, D.J.; Bloomer, S.H.; MacLeod, C.J.; Taylor, B.; Goodlife, A.M. Bathymetry of the Tonga Trench and Forearc: A map series. *Mar. Geophys. Res.* **2000**, *21*, 489–512. [[CrossRef](#)]
30. Cronin, S.; Brenna, M.; Smith, I.; Barker, S.; Tost, M.; Ford, M.; Tonga’onevai, S.; Kula, T.; Vaioumunga, R. New Volcanic Island Unveils Explosive Past. *Eos* **2017**, *98*, 1. [[CrossRef](#)]
31. Kosyakov, S.I.; Kulichkov, S.N.; Chunchuzov, I.P. Distinctive Features of the Development of the Tonga Underwater Volcano Eruption According to Acoustic Monitoring Data. *Pure Appl. Geophys.* **2025**, *182*, 2277–2290. [[CrossRef](#)]
32. Fujii, Y.; Satake, K. Modeling the 2022 Tonga Eruption Tsunami Recorded on Ocean Bottom Pressure and Tide Gauges Around the Pacific. *Pure Appl. Geophys.* **2024**, *181*, 1793–1809. [[CrossRef](#)]
33. Plank, S.; Ciancia, E.; Genzano, N.; Falconieri, A.; Martinis, S.; Taubenböck, H.; Pergola, N.; Marchese, F. The evolution of the 2022–2024 eruption at Home Reef, Tonga, analyzed from space shows vent migration due to erosion. *Sci. Rep.* **2025**, *15*, 11508. [[CrossRef](#)]
34. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [[CrossRef](#)]
35. 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. [[CrossRef](#)]
36. Percacci, R. *Dataset for Land/Water Semantic Segmentation in Tonga and Other Pacific Regions*; Elsevier: Amsterdam, The Netherlands, 2025. [[CrossRef](#)]
37. Ba, R.; Chen, C.; Yuan, J.; Song, W.; Lo, S. SmokeNet: Satellite Smoke Scene Detection Using Convolutional Neural Network with Spatial and Channel-Wise Attention. *Remote Sens.* **2019**, *11*, 1702. [[CrossRef](#)]
38. Chen, S.; Cao, Y.; Feng, X.; Lu, X. Global2Salient: Self-adaptive feature aggregation for remote sensing smoke detection. *Neurocomputing* **2021**, *466*, 202–220. [[CrossRef](#)]
39. Müller, D.; Soto-Rey, I.; Kramer, F. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Res. Notes* **2022**, *15*, 210. [[CrossRef](#)]
40. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity Mappings in Deep Residual Networks. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part IV 14; pp. 630–645. [[CrossRef](#)]
41. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778. [[CrossRef](#)]
42. Chollet, F. Keras. 2015. Available online: <https://github.com/fchollet/keras> (accessed on 1 July 2024).
43. Iakubovskii, P. Segmentation Models. 2019. Available online: https://github.com/qubvel/segmentation_models (accessed on 1 July 2024).
44. Sha, Y. Keras-Unet-Collection. 2021. Available online: <https://github.com/yingkaisha/keras-unet-collection> (accessed on 1 July 2024).
45. Simonyan, K.; Vedaldi, A.; Zisserman, A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *arXiv* **2014**, arXiv:1312.6034. [[CrossRef](#)]
46. Montavon, G.; Binder, A.; Lapuschkin, S.; Samek, W.; Müller, K.R., Layer-Wise Relevance Propagation: An Overview. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*; Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., Müller, K.R., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 193–209. [[CrossRef](#)]
47. Yeo, I.A.; McIntosh, I.M.; Bryan, S.E.; Tani, K.; Dunbabin, M.; Metz, D.; Collins, P.C.; Stone, K.; Manu, M.S. The 2019–2020 volcanic eruption of Late’iki (Metis Shoal), Tonga. *Sci. Rep.* **2022**, *12*, 7468. [[CrossRef](#)]
48. Global Volcanism Program. Report on Lateiki (Tonga) (Crafford, A.E. and Venzke, E., Eds.). Bulletin, Global Volcanism Network, 45:2, Smithsonian Institution. 2020. Available online: <https://volcano.si.edu/showreport.cfm?doi=10.5479/si.GVP.BGVN202002-243070> (accessed on 22 January 2025).
49. Global Volcanism Program. Report on Home Reef (Tonga) (Sennert, S., Ed.). Weekly Volcanic Activity Report, 11 October–17 October 2023. Smithsonian Institution and US Geological Survey. 2023. Available online: <https://volcano.si.edu/showreport.cfm?wvar=GVP.WVAR20231011-243080> (accessed on 23 January 2025).

50. Global Volcanism Program. Report on Home Reef (Tonga) (Sennert, S., Ed.). Weekly Volcanic Activity Report, 3 July–9 July 2024. Smithsonian Institution and US Geological Survey. 2024. Available online: <https://volcano.si.edu/showreport.cfm?wvar=GVP.WVAR20240703-243080> (accessed on 5 September 2025).
51. Global Volcanism Program. Report on Home Reef (Tonga) (Sennert, S., Ed.). Weekly Volcanic Activity Report, 11 December–17 December 2024. Smithsonian Institution and US Geological Survey. 2024. Available online: <https://volcano.si.edu/showreport.cfm?wvar=GVP.WVAR20241211-243080> (accessed on 23 January 2025).
52. Terry, J.P.; Goff, J.; Winspear, N.; Bongolan, V.P.; Fisher, S. Tonga volcanic eruption and tsunami, January 2022: Globally the most significant opportunity to observe an explosive and tsunamigenic submarine eruption since AD 1883 Krakatau. *Geosci. Lett.* **2022**, *9*, 24. [[CrossRef](#)]
53. Wieland, M.; Fichtner, F.; Martinis, S.; Groth, S.; Krullikowski, C.; Plank, S.; Motagh, M. S1S2-Water: A global dataset for semantic segmentation of water bodies from Sentinel-1 and Sentinel-2 satellite images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *17*, 1084–1099. [[CrossRef](#)]
54. Li, Y.; Dang, B.; Li, W.; Zhang, Y. GLH-Water: A Large-Scale Dataset for Global Surface Water Detection in Large-Size Very-High-Resolution Satellite Imagery. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; Volume 38, pp. 22213–22221. [[CrossRef](#)]

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.