Article

Chimney Identification Tool for Automated Detection of Hydrothermal Chimneys from High-Resolution Bathymetry Using Machine Learning

Isaac Keohane * and Scott White

School of the Earth, Ocean and Environment, University of South Carolina, Columbia, SC 29208, USA;
swhite@geol.sc.edu
* Correspondence: ikeohane@seoe.sc.edu

Abstract: Identifying the locations of hydrothermal chimneys across mapped areas of seafloor spreading ridges unlocks the ability to research questions about their correlations to geology, the cooling of the lithosphere, and deep-sea biogeography. We developed a Chimney Identification Tool (CIT) that utilizes a Convolutional Neural Network (CNN) to classify 1 m gridded AUV bathymetry and identify the locations of hydrothermal vent chimneys. A CNN is a type of Machine-Learning model that is able to classify raster data based on the shapes and textures in the input, making it ideal for this task. The criteria that have been used in previous manual classifications of chimneys have focused on the round base and spire shape of the features, and are not easily quantifiable. Machine-Learning techniques have previously been implemented with sonar data to classify seafloor geology, but this is the first application of these methods to hydrothermal systems. In developing the CIT, we compiled the bathymetry data from two rasters from the Endeavor Ridge—each gridded at a 1 m resolution—containing 34 locations of known hydrothermal chimneys, and from the 92° W segment of the Galapagos Spreading Center (GSC) containing 14. The CIT produced a primary group of outputs with 96% agreement with the manual classification; moreover, it correctly caught 29 of the 34 known chimneys from Endeavor and 10 of the 14 from the GSC. The CIT is trained to identify features with the characteristic shape of a hydrothermal vent chimney; therefore, it is susceptible to the misclassification of unusually shaped cases, given the limited training data. As a result, to provide the option of having a more inclusive application, the CIT also produced a secondary group of output locations with 61% agreement with the manual classification; moreover, it caught three of the four additional known chimneys from the GSC and four of the five from Endeavor. The CIT will be used in future investigations where an inventory of individual chimneys is important, such as the cataloguing of off-axis hydrothermal venting and the investigation of chimney distribution in connection to seafloor eruptions.

Keywords: Machine Learning; hydrothermal venting; AUV; remote sensing; classification; bathymetry; Neural Networks

1. Introduction

High-resolution mapping using Autonomous Underwater Vehicles (AUVs) at seafloor spreading ridges has ushered in new capabilities for investigating hydrothermal venting at the individual-vent scale. Without meter-scale mapping, hydrothermal vents have typically been visually identified using manned and remotely operated submersibles that investigate areas where water-column temperature or particle anomalies exist [1–3]. These surveys focus primarily on active vents and require the submersibles to operate within visual range of the seafloor, at <10 m; this restricts their speed and, thus, the area explored. However, AUV multibeam mapping surveys conducted ~50 m off-bottom produce bathymetry that can be gridded at a ≤1 m resolution [4]. These maps can cover tens to hundreds of square kilometers, while also resolving active and inactive hydrothermal chimneys and mounds

that are only a few meters wide and high (Figure 1) [4–6]. Identifying the locations of hydrothermal chimneys across mapped areas of seafloor spreading ridges unlocks the ability to investigate questions about their spatial distributions and correlations to related characteristics of ridge geology, and ultimately, larger questions about the flux of seawater and seawater composition. For example, Jamieson et al. [5] estimated the volume of hydrothermal sulfide edifices over a 62 km^2 area of the Endeavor segment of the Juan de Fuca ridge by identifying active and inactive chimneys resolved by AUV bathymetry. Clague et al. [7] observed the clustering of hydrothermal chimneys within segments along the ridge axis of the Endeavor segment, and compared the density of chimneys to the observed distribution from the Alarcón rise identified by Paduan et al. [6].
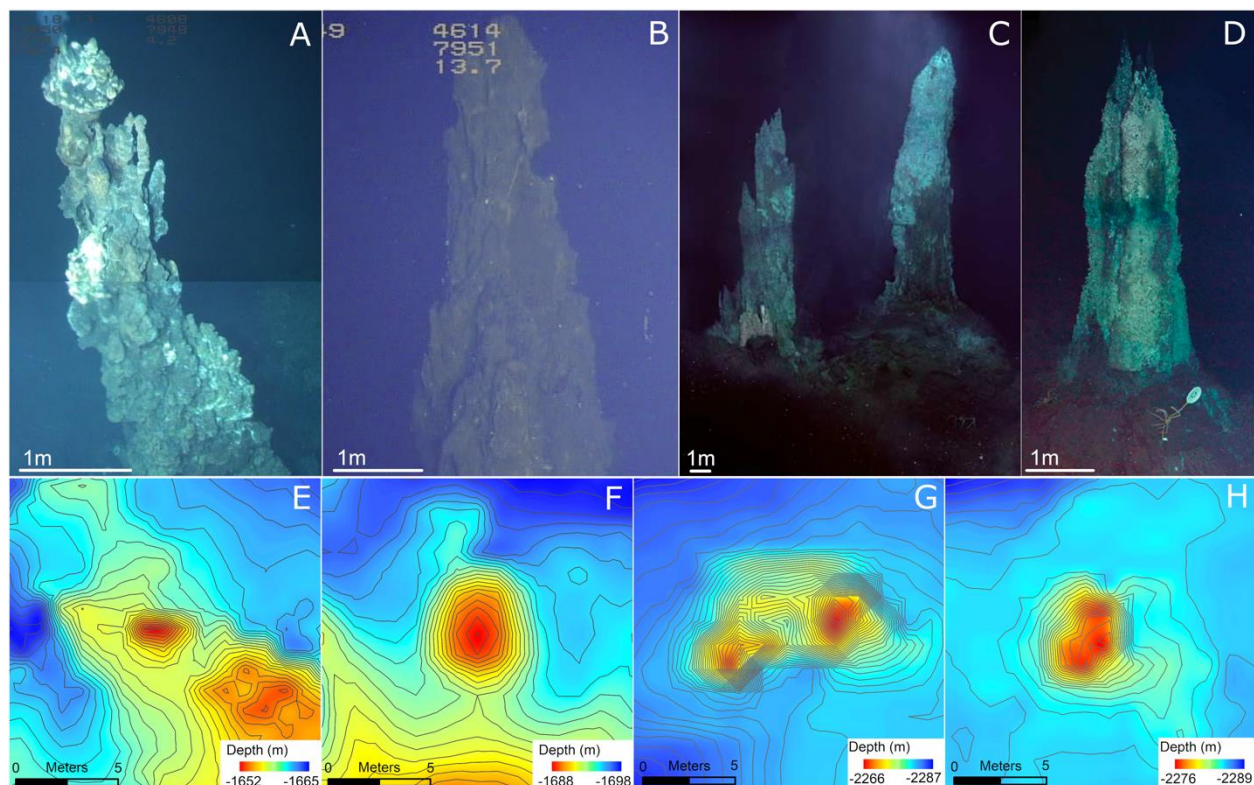
Figure 1. Photographs of hydrothermal chimneys (**A**–**D**) and their expression in bathymetry, gridded at 1 m with 0.5 m contour lines (**E**–**H**, respectively). Because of parallax, white scale bars in (**A**–**D**) represent estimations for scale at the chimney location in the image. (**A**,**E**) and (**B**,**F**) are examples from the GSC (Expedition AT15-63, J. Sinton, chief sci.). (**A**) is a mosaic of 2 images and shows the entire chimney edifice, and (**B**) shows only the top half of the chimney. (**C**,**G**) show 2 chimneys from the Stonehenge field at Endeavor, and (**D**,**H**) show the Bridgit vent. Photographs modified from Glickson et al. [3].

High-resolution AUV bathymetry exists at several seafloor spreading ridges and is rapidly becoming much more available. However, classifications of these datasets to identify hydrothermal chimneys from bathymetry have exclusively been performed manually, with individuals looking at visualizations of the bathymetry and selecting features that resemble known chimneys [5–9]. In this study, we begin to develop criteria for identifying chimneys from their characteristic spire shape in bathymetry (Figure 1). Previous manual picking of hydrothermal chimneys developed a general consensus that chimneys manifest in 1 m gridded bathymetry as steep-sided spires, several meters wide, with a round to semi-round base, and have a height range of roughly 3–30 m. Identifying chimneys by manual selection requires substantial time and effort and is not strictly reproducible, as the interpretation is affected by the skill and attention of the individual performing the

classification, which varies between people and for an individual over time. An automated computer tool that algorithmically searches high-resolution bathymetry and identifies the hydrothermal chimneys solves both of these issues.

We developed a Chimney Identification Tool (CIT) that utilizes a Convolutional Neural Network (CNN), which works with 1 m gridded bathymetry to identify the locations of hydrothermal vent chimneys. A CNN is a type of Machine-Learning model that classifies raster data by building a network of relative connections between the pixels within the raster [10,11]. The nested nature of these connections allows classification based not only on the difference in input pixel values, but also on the incredibly complex relationships between pixels arranged across different parts of the raster that humans visually interpret as shape or texture [11–13]. The CNN is uniquely useful in this context of hydrothermal chimney detection where the shapes of features, and not just specific raster values, are a key part of the criteria that have been used in previous manual classifications. One drawback of a CNN is that due to the abstract mathematical structure and complexity, it operates as a black box; it can be successfully trained to classify input, but what it is actually "seeing" in the input is impossible to extract and examine in human terms [14,15]. Machine-Learning techniques have been successfully implemented with sonar data to classify seafloor geology in many studies (e.g., [16–20]); however, to our knowledge, this is its first application in hydrothermal chimney identification.
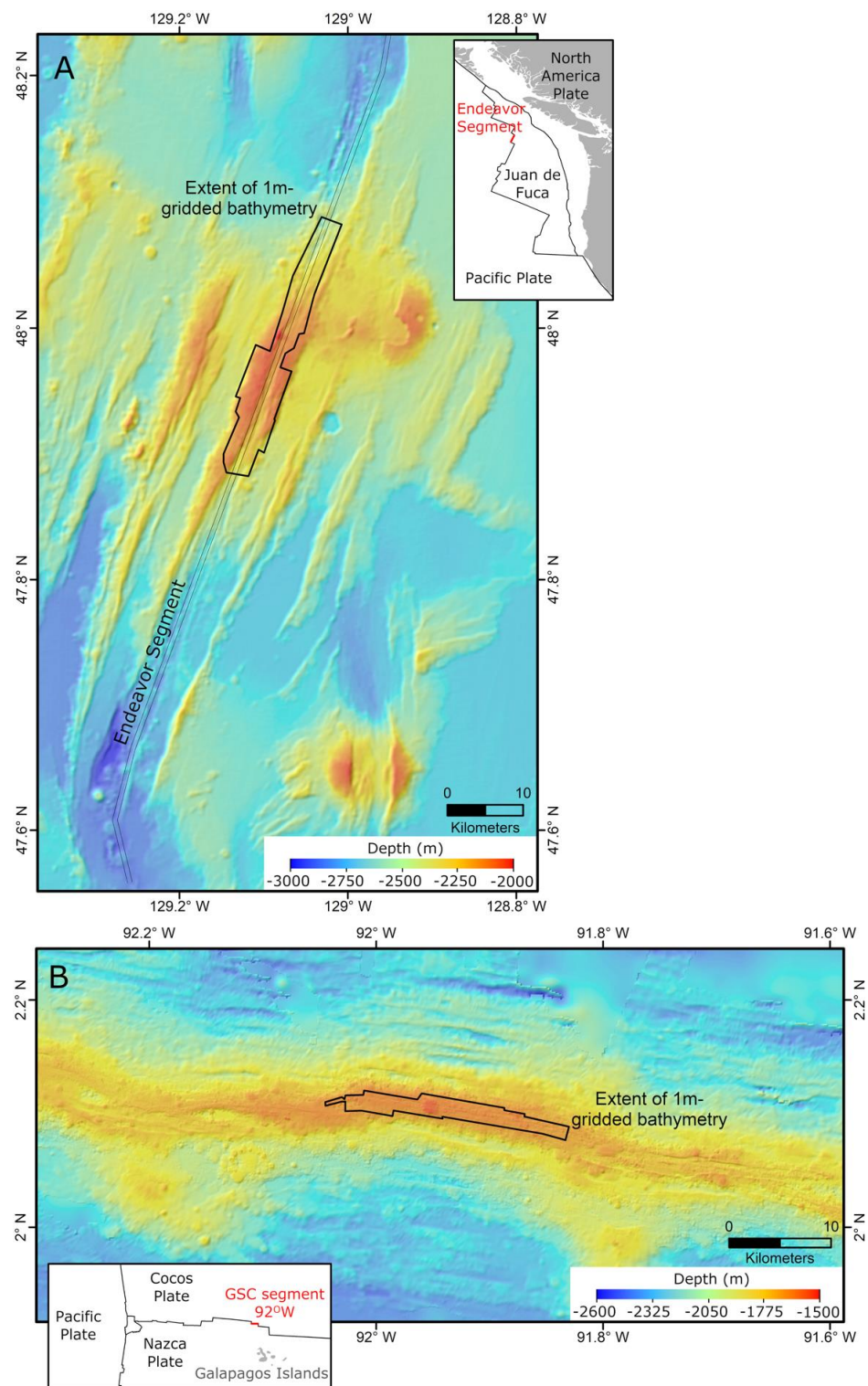
2. Study Areas

The CIT was developed using data from two study areas: the Endeavor segment of the Juan de Fuca Ridge and the 92° W segment of the Galapagos Spreading Center (GSC) (Figure 2). These were chosen not only because processed AUV multibeam bathymetry is available at meter-scale resolution [7,18], but also because previous manual classifications to identify hydrothermal chimneys have been conducted on both bathymetry datasets [7,8] (Figures 3 and 4). These inventories of chimney locations provide a ground-reference dataset that is necessary for training a supervised automated classifier, and also provide pre-existing testing data from completely independent manual classification.

The 92° W segment of the GSC and the Endeavor Segment have similar average spreading rates of 54 and 52 mm/year, respectively [21], and provide a valuable comparison because they have very different geology and bathymetry. The GSC 92° W segment has high levels of magma supply due to its proximity to the Galapagos hotspot; this results in a shallow axial high (1600 m), much less faulting than at the Endeavor Segment, and the presence of unfaulted volcanic cones and mounds (Figures 2 and 4) [22,23]. The Endeavor Segment has a large axial graben 0.8–1.7 km wide and 100–200 m deep, and a much smaller area covered by unfractured lava flows (Figures 2 and 3) [24,25]. These geologic environments directly relate to the process of developing the CIT to work in a range of geologic environments; this is because Endeavor has many more steep-sided features and more fractured terrain than GSC, and contains numerous large chimneys due to the combination of increased permeability from the fracturing and the lack of volcanic resurfacing [7]. Developing a CIT to effectively identify chimneys in both of these settings, with characteristics of both low and high magma-supply environments, suggests that it has the capability to be deployed in a wide range of seafloor spreading ridge environments.

Before developing the CIT, we compiled the bathymetry data from each area into two rasters, each gridded at a 1m-resolution. The Clague et al. [7] dataset of manual picks from Endeavor contains 406 locations of hydrothermal chimneys (Figure 3), and the White and Lee [8] dataset from the GSC contains 122 (Figure 4). There is a subset of each set of manual pick locations that are known vents and are confirmed by direct observation with a submersible. In the Clague et al. [7] Endeavor dataset, these examples were identified as those with individual chimney names, and in the White and Lee [8] GSC dataset, these confirmed chimneys are un-named. There are 34 confirmed chimneys in the Endeavor dataset and 14 for the GSC.

Figure 2. Regional bathymetry for the 2 study areas: (**A**) the Endeavor Segment of the Juan de Fuca Ridge; (**B**) the 92° W segment of the Galapagos Spreading Center (GSC). The extents of the 1 m gridded bathymetry rasters are outlined. Insets show the plate boundaries and plate names wherein red lines indicate the specific segment of the plate boundary in which the study areas are located. Basemap bathymetry is from GMRT [26].
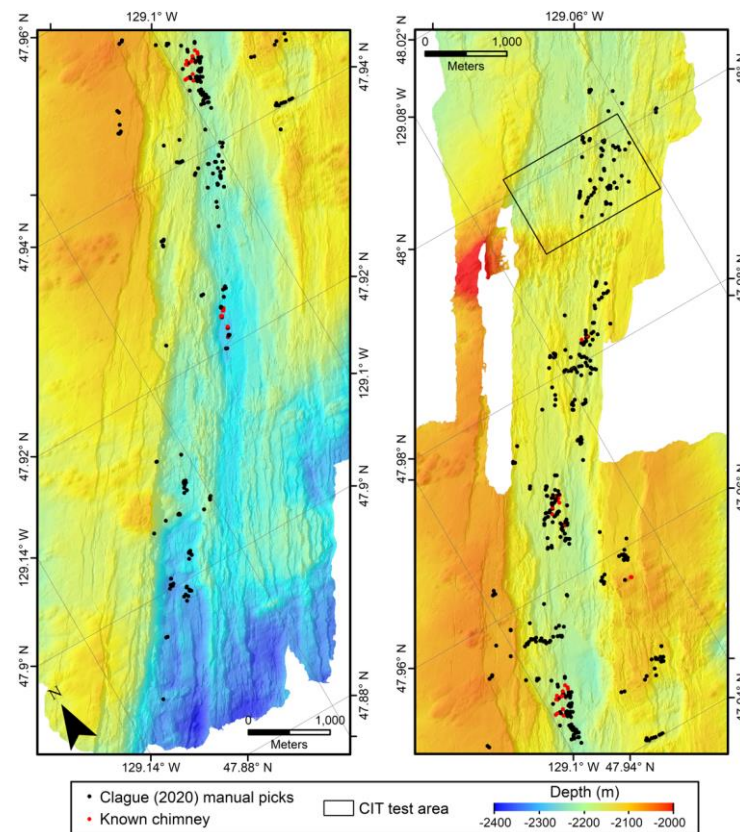
Figure 3. Shaded relief bathymetry map of the Clague et al. [7] manually selected chimney locations (black dots) for the Endeavor study area's 1 m gridded AUV bathymetry. The black box outlines the test area that was used to evaluate the CIT in this study. Note the clustering of chimney locations, the pronounced axial graben, and pervasive faulting, expressed as NE-SW trending lineaments, in the bathymetry.
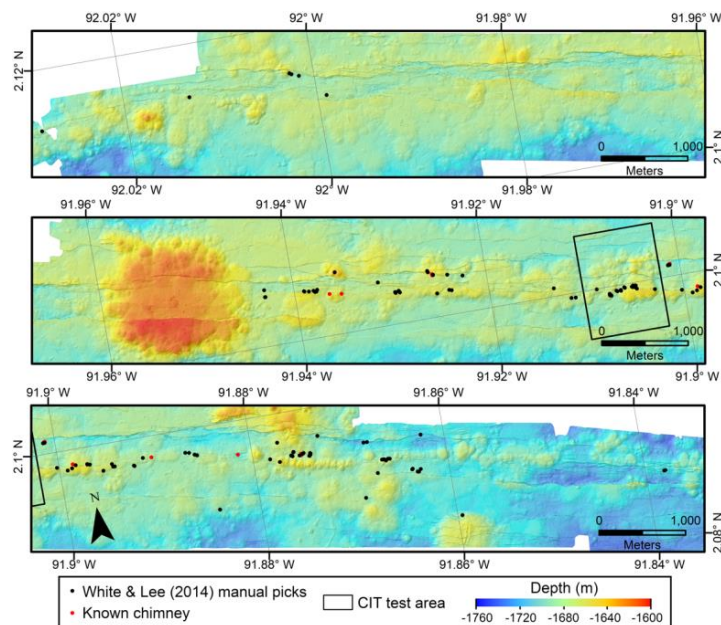
Figure 4. Shaded relief map of the 1 m gridded AUV bathymetry of the GSC study areas [18], with the White and Lee [8] manually picked chimney locations plotted as black dots. The black box outlines the test area that was used to evaluate the CIT. Note the pronounced volcanic cone and numerous volcanic mounds, along with much lower expression of faulting compared to the Endeavor Ridge segment in Figure 3.

3. Methods

The initial construction of the CIT came from implementing techniques from the field of image classification. The idea of using a selective search to classify subsets of a larger raster using a Neural Network is known in the field of computer vision as Object Detection [27]. We drew direct guidance on how to apply this technique—originally developed for analyzing digital photographs—to bathymetric rasters from Bycroft et al. [28], who adapted it for high-resolution surveying with unmanned aerial vehicles, and Valentine et al. [20], who classified seafloor bathymetry to identify seamounts. Bycroft et al. [28] found a CNN successful at identifying round crab burrows which have a footprint of typically <15 pixels in diameter; this is a similar pixel footprint to hydrothermal chimneys in 1 m gridded bathymetry. Valentine et al. [20] used a selective search to extract subset patches of the larger bathymetric raster and identified discrete output locations by identifying extrema in a model output raster, inspiring the use of similar techniques in the CIT. The overall structure of the CIT workflow was derived and modified from these previous applications (Figure 5).

We refined the CIT workflow to tailor it specifically to the task of hydrothermal chimney detection through iterative testing; we achieved this by utilizing two criteria for evaluation, each focused on a different type of classification error, omission, or commission. Omission was measured by running each known (visually observed) hydrothermal chimney from both the Endeavor and GSC datasets (34 and 14, respectively) through the CIT and confirming whether they were correctly identified. The number of these known chimneys that were correctly identified gave a measure of how effectively the CIT met the target goal of identifying all the chimneys. Commission was evaluated by generating a CIT output for a test area raster from each study area (areas outlined in Figures 3 and 4). We manually classified each of the generated locations, plotting the bathymetry and 0.5m contours in ArcGIS, and positively labeling round features with $\geq$5 closed 0.5 m contours that resembled the size and shape of known chimneys. We studied both the bathymetry associated with the known hydrothermal chimneys and the methods by which previous manual classifications were conducted, to gain an understanding of how hydrothermal chimneys look in bathymetry: steep-sided spires several meters wide, with a rounded footprint and a height range of roughly 3 to 30 m tall. The number of CIT output locations that also passed our manual classification gave a measure of how effectively the CIT included only chimneys in the output. Throughout the development process, we were often faced with a choice between optimizing for lower omission or lower commission. In these cases, we chose to optimize for lower omission, wanting to ensure that the CIT had the lowest chance possible of missing hydrothermal vent chimneys if they were present in the input data.

The CIT was built using Python 3.7 and PyTorch 1.6 [29]. A GitHub repository containing the Python scripts created for this project is linked in the data availability statement. The initial stage of raster filtering was not automated within the CIT, and was performed manually using a combination of ArcGIS Desktop 10.7, R 4.0.3, and Python 3.7.

3.1. Filtering Bathymetry to Produce Multiple Derived Rasters

The CIT was developed to take 3 raster layers as inputs—(1) normalized bathymetry, (2) normalized local slope, and (3) normalized Bathymetric Position Index (BPI)—derived from an initial bathymetric raster (Figure 6). During the development process, we found that utilizing this combination of 3 layers resulted in consistent improvements in omission and commission errors compared to using a single layer. Pixel values in the derived rasters are the outputs of a function applied to a square moving-window neighborhood of pixels. Each raster layer is normalized so that all pixel values are within the range of −1 to 1. This is a standard step when using a CNN as it brings the input values to a similar magnitude as the function coefficients contained in each node of the network.
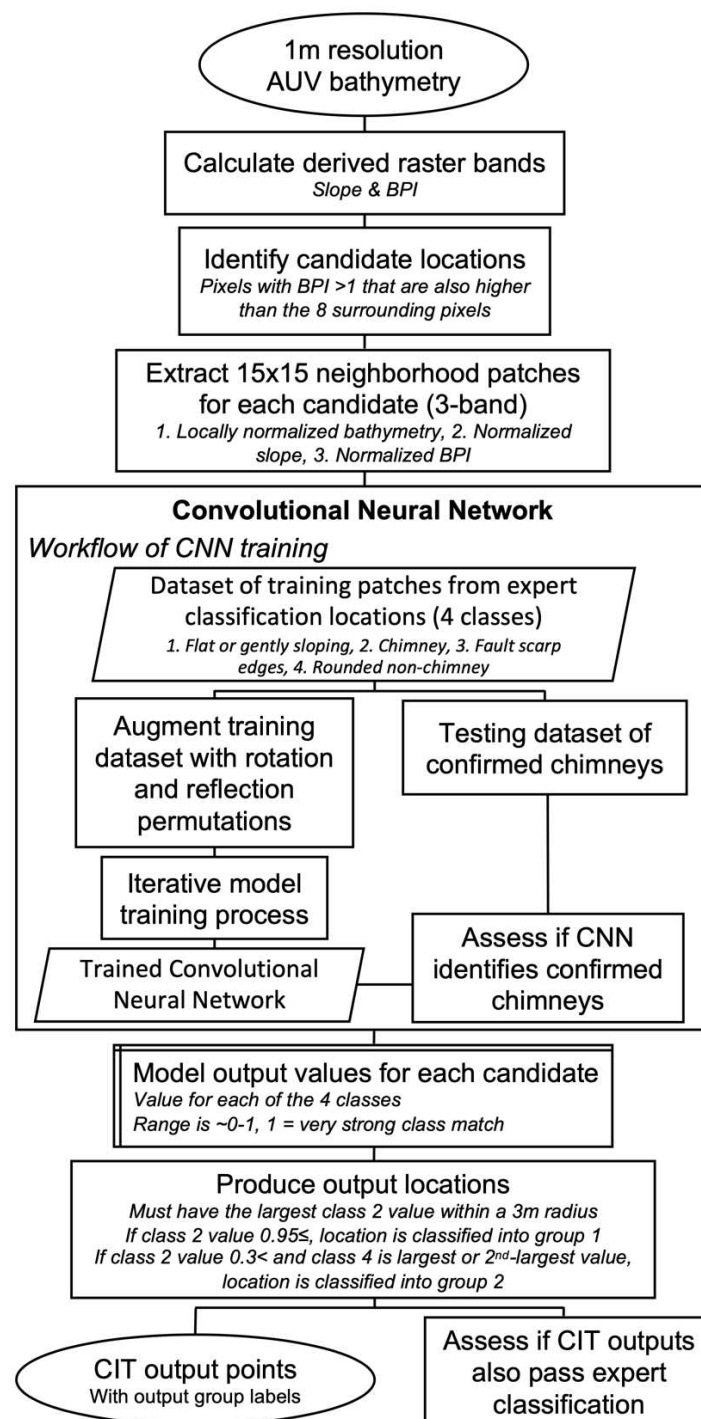
Figure 5. CIT workflow to generate output chimney locations from input 1 m gridded bathymetry. Contained in the box representing the Convolutional Neural Network is a sub-flowchart for the model training process.
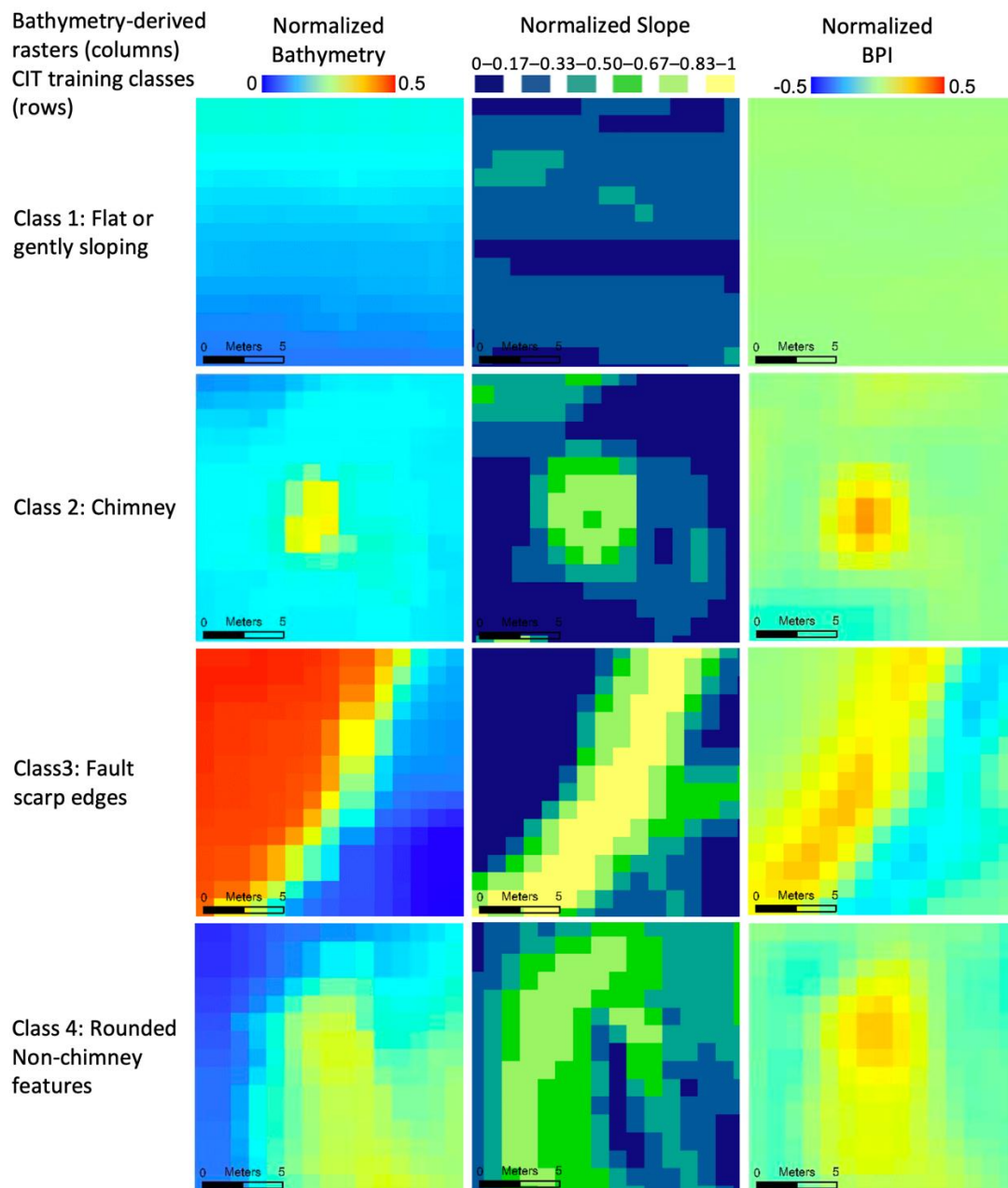
Figure 6. Examples of the 3 bathymetry-derived rasters (columns) that are sent into the CNN as 15 × 15m (15 × 15 pixel) patches, as described in the text. Each row represents a typical example from each of 4 training classes.

The first raster layer is a locally normalized bathymetry. The textures and shapes observed in this layer capture the seafloor surface just as the raw bathymetry does, but it can be directly compared between different areas that vary in absolute water depth. Each pixel value is scaled by subtracting the minimum value within a 30 m wide neighborhood window, and then dividing it by a scaling factor of 50 m. This 50 m scaling factor was chosen to be greater than the largest range of bathymetry, which was observed to be within a 30 m neighborhood from the GSC and Endeavor datasets. In future applications of the CIT, using this constant scaling factor could potentially lead to pixels from an area with larger variation than was observed in these 2 study areas, having pixel values greater than 1. This would not directly lead to the CIT being ineffective, as a CNN can take input values outside of the −1 to 1 range; however, if these input values were several times larger than 1, it could negatively affect performance. CIT is developed to handle inputs

from different bathymetry datasets and areas. If each input was normalized using the minimum and maximum values of its containing raster, features of the same size and shape from different areas would not translate in the same way when normalized. Choosing this constant scaling factor is necessary to ensure that examples from different rasters are normalized consistently.

The second layer is the local slope of median-filtered bathymetry using a 3 × 3 pixel (or 3 m square) neighborhood. This layer represents the rate of change of the bathymetry and is calculated by taking the maximum change in depth between each pixel and its 8 neighbors. This slope can be calculated as an angle relative to a flat surface of constant bathymetry. With calculating slope as an angle, the resulting layer can be normalized by simply dividing by 90°, as that is the largest possible value.

The third layer is the normalized Bathymetric Position Index (BPI), with an inner neighborhood of 3 m and an outer neighborhood 11m in diameter [30]. BPI calculates the difference in the mean bathymetry between the inner and outer neighborhoods. At the meter scale, features such as chimneys and the upper edges of fault scarps show up as more positive values, and features such as pits, fissures, and the lower edges of fault scarps show up as more negative values (Figure 6). Flat areas, including those that slope uniformly over 10 m or larger spatial scales, have BPI values close to zero (Figure 6). This layer was normalized similarly to the locally normalized bathymetry by using a scaling value of 20, resulting in the same considerations for future applications that might have BPI values larger than 20. This value was chosen to be greater than largest absolute value observed in any of the input training areas (18.0). The mean for BPI was ~0, with positive and negative values indicating local highs and lows, respectively; therefore, dividing by this constant value rescaled all of the input values to within the desired range of −1 to 1, while preserving the sign. Layers 1 and 2 do not contain negative values, because there is no change in sign associated with the un-normalized raster layers as there is with BPI. This difference in the range of values between layers is not a problem for the CNN, as this normalization step is needed to adjust the absolute values of inputs to be a similar magnitude to the internal coefficients.

3.2. Selective Search to Extract Candidate Patches

In order to improve the runtime of the CIT, a simple logical filter is used to identify potential chimney candidate locations from the entire study area, and then local 15 × 15-pixel patches are extracted for input to the CNN, only from those areas that passed this preliminary set of criteria. The goal of this filter is to rapidly restrict the search by eliminating any areas that contain no chimney-like features. Pixels are selected by this preliminary search if they are local maxima in the bathymetry compared to the 8 adjacent pixels and have a BPI value ≥1. The BPI values used were from the 3 m/11 m BPI raster used as input into the CNN, but before the normalization was applied. This threshold value of 1 for the filter value was selected to be extremely conservative based on our empirical observation of the training data. The spire of a hydrothermal chimney will always be a local maximum in 1 m gridded bathymetry, and no hydrothermal chimneys in the two training datasets have an observed BPI value below 1. For each candidate pixel that passed the selective search, a 15 × 15 raster patch centered on that pixel is extracted. The 3-band normalized raster for each patch is sent into a CNN for classification. This logical filter does successfully execute the role of restricting the local patches needed for the search, as less than 1% of the pixels from the input raster are usually passed.

3.3. Evaluating Each Candidate with a CNN

The CNN produces output values for 4 different classes that it was constructed to classify between: (1) flat or gently sloping areas, (2) chimneys, (3) fault scarp edges, and (4) rounded non-chimney features (Figure 7). Each of these output values ranges roughly from 0 to 1, with 1 being a strong class match. Additionally, the sum of all 4 for a given

output is roughly equal to 1, so equal confusion in the CNN between 2 classes would result in each of those classes having an output value of ~0.5.
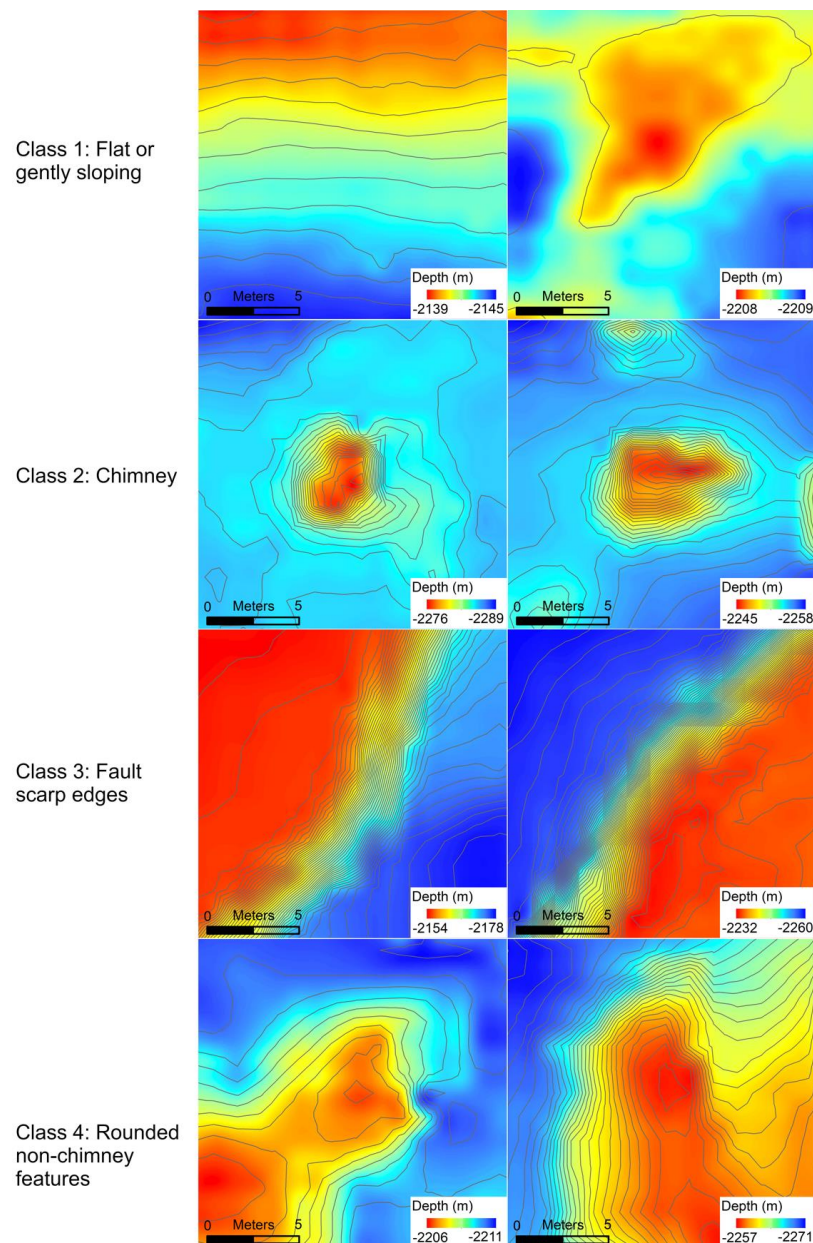
Figure 7. Bathymetry maps showing examples of the features picked for the classes used for training the CNN. Two examples of each class are provided. Each panel is 15 × 15 m, which is the size of each patch sent into the CNN. Grey lines are 0.5 m contours.

Before implementation in the CIT, the CNN was trained with reference examples from the 4 classes. These examples were 15 × 15 raster patches of the 3 normalized raster layers extracted for each point in a set of training locations. Training examples for classes 1, 3, and 4 were originally produced for this project by manually identifying locations in the two study areas. We generated 300 of each class from the Endeavor dataset and 140 of class 1 and 3 from the GSC dataset. We only generated 85 examples for class 4 from the GSC dataset, as the frequency of these rounded non-chimney features was much lower than at Endeavor. The chimney (class 2) examples were adapted from the existing manual classification picks of Clague et al. [7] and White and Lee [8]. We adjusted these locations, by no more than 5 m, so that the center pixel of each training patch would be the exact peak

of the chimney. Ensuring that each chimney training example has the peak in the center minimizes unnecessary variation that could confuse the CNN and leads to a model that is trained to identify patches that not only contain a chimney, but are specifically centered on one. This ties directly into the end goal of the CIT, which is to identify a distinct point location for each chimney.

One of the common difficulties with training Machine-Learning models is the limited number of training examples. Between the two study areas, there were only ~500 chimney examples available for training the model. We were only able to reach the model performance of this version of the CIT after generating artificial training examples by transforming the original ones [31]. Before model training, new permutations of each 15 × 15 m example patch from the 4 classes are generated using all 8 unique combinations of 90° rotations and mirror flips. Chimneys are theoretically radially symmetric features and, therefore, can be viewed as transformation invariant for rotations and flips. Creating these new examples from modifications of the existing ones has the effect of multiplying the training dataset 7×. After augmentation of the training dataset, the original training examples for the 48 known chimneys, as well as those located in both test areas, were removed from the training dataset, as these examples are used to evaluate the CIT performance.

The CNN model is trained by iteratively providing single 15 × 15 example patches of a known class, running the CNN model, evaluating the output against that known class, then updating the values of the nodes within the network in a way that would improve the classification of that example. Because this is a very basic overview, an in-depth explanation of the mathematical theory behind the structure of a Convolutional Neural Network and the iterative training process is beyond the scope of this paper; however, it is extensively covered in other literature [10–13,31]. The order in which these examples are shown to the network is randomized, and the set of training examples is sampled so that the CNN sees the same total number of examples from each class by the end of the training process. The CIT uses a mean-square-error loss for evaluating the accuracy of each training iteration, and a stochastic gradient descent method for determining how the nodes should be updated [32,33]. The amount of change applied to the CNN on a single iteration is controlled by a learning-rate parameter. The CNN in the CIT is trained using epochs. An epoch is a full cycle of training whereby the set of training examples is sent through the model. Between each Epoch, the order of the training examples is rerandomized and the learning rate incrementally decreases. The model training ends after the model performance converges using the criteria, which state that the average mean-square-error loss for the epoch is <0.005 and that the value has changed by less than 5% from the previous epoch.

3.4. Deriving Distinct Chimney Locations from the Output Values of the CNN

The CIT generates chimney locations from the candidate pixel locations by evaluating the output values produced by the CNN. In order to be classified as a chimney, the CNN output value for the chimney class (class 2) must be higher than any other chimney (class 2) value within a 3 m radius. Any two candidate pixels that are within this distance are assumed to be from the same feature. From these, two sets of classified chimney locations are generated. First, pixels are assigned to a 'most likely chimney' set if the class 2 value is above 0.95, as there is a distinct peak of examples with values near 1 and above a threshold near 0.95 (Figure 8).
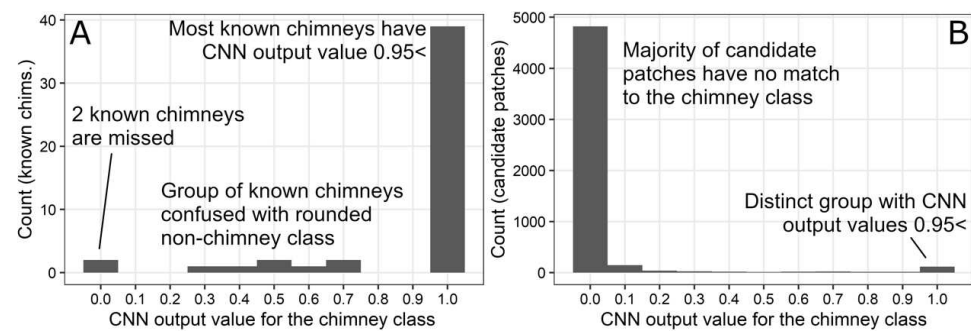
Figure 8. Annotated histograms of the chimney class (class 2) CNN output values for (**A**) the set of known chimney examples and (**B**) the set of candidate patches extracted from the two test areas. The distinct population of output values above 0.95 are classified as the 'most likely chimneys' group. The 7 examples of known chimneys with intermediate class 2 values are caught by the 'probable chimneys' in the CIT output.

To provide the ability for the CIT to be as inclusive of hydrothermal chimneys as possible, the CIT generated a second 'probable chimney' set of points with a combination of intermediate class 2 and class 4 values. When we evaluated how the CNN classified the examples of known chimneys, we observed a smaller but distinct group with class 2 values of 0.3–0.75 (Figure 8). Each of these examples had a corresponding class 4 value for which the sum of both was equal to ~1. We interpreted this group of examples to be chimneys (class 2) that the model was unable to fully discern from the features of class 4, which loosely resembled chimneys. To allow for the inclusion of chimneys that might be confused with class 4, the CIT generated this second group of 'probable chimney' outputs, with class 2 CNN output values in the range of 0.30–0.95, and for which the sum of class 2 and class 4 values was $\geq$0.95.

4. Results

Looking qualitatively at the CIT output for the Main Endeavor Field suggests that the CIT consistently identifies local peaks that resemble chimneys and identifies all of the known (named) chimneys (Figure 9). The CIT output is very similar to the chimney picks from Clague et al. [7]. All of the identified features are rounded, spiky features, and no outputs are erroneously placed on features such as fault scarp edges that obviously are not chimney-like.
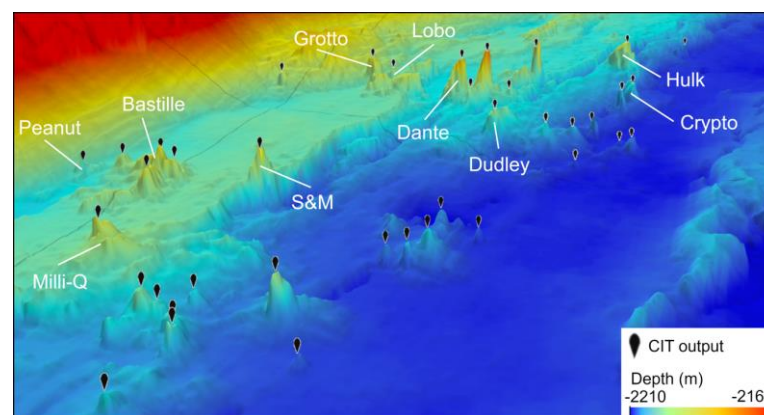
Figure 9. Perspective plot of 1 m gridded bathymetry of the Main Endeavor Field from the Endeavor segment. Black flags indicate the 'most likely chimney' locations identified by the CIT. White text indicates named hydrothermal chimneys that have been directly confirmed by submersible dives. The distance between Hulk and Milli-Q is 342 m. Note how the CIT identifies all the confirmed chimneys and only distinct spire-shaped features, which are likely un-named chimneys because they are either inactive or not visited by a submersible.

The CIT correctly classified as 'most likely chimneys' 29/34 known chimneys from Endeavor and 10/14 from the GSC, for a combined accuracy of 81.3% (Table 1). The CIT classified 95.8% of known chimneys when the more inclusive set of 'probable chimneys' was also included. We note that the two known chimneys that were missed by the CIT are very small, each only consisting of one bathymetry pixel that rises significantly above the surrounding area (Figure 10).

Table 1. Assessment of known chimneys caught by the CIT.

Study Area	Missed	'Probable Chimneys'	'Most Likely Chimneys'	Total	'Most Likely Chimneys' %	Either Group %
GSC	1	3	10	14	71.4%	92.9%
Endeavor	1	4	29	34	85.3%	97.1%
Combined	2	7	39	48	81.3%	95.8%
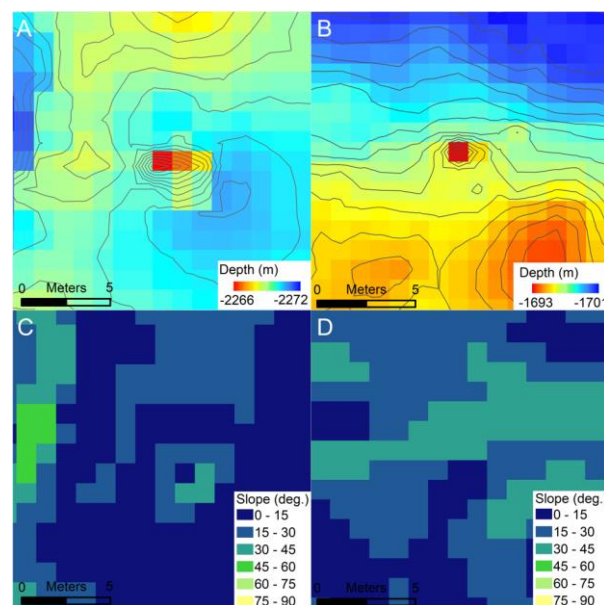
Figure 10. Bathymetric patches (**A**,**B**) of two known chimneys that were completely missed by the CIT; in other words, the CNN output value for the chimney class was roughly 0. At this scale, individual pixels are easily seen; each of these chimneys is defined almost exclusively by only 1 pixel higher than the surrounding area. The slope of median filtered bathymetry, before normalization for use in the CIT is shown in (**C**,**D**), respectively. The chimney shown in (**A**,**C**) is from Endeavor, and the one in (**B**,**D**) is from the GSC. The smoothing of these single pixels is likely a major factor in their misclassification, and illustrates the lower size limit of detection in 1 m gridded bathymetry data.

The CIT identified 78 'most likely chimneys' within the two test areas, with 63 features from Endeavor and 15 from the GSC (Table 2). Compared to our manual chimney picks of those that strongly resembled, and were interpreted to be, hydrothermal chimneys, 61/63 and 14/15 matched from the Endeavor and GSC test areas, respectively. This translates to 96% agreement across both areas combined. The less selective 'probable chimney' set did introduce more commission error, as anticipated, matching 38/61 and 5/10 manual picks from the Endeavor and GSC test areas, respectively, for a combined accuracy of 61% (Table 2).

Table 2. Assessment of the CIT outputs from the test areas that also passed manual classification criteria.

| Study Area	Output Group	Passed	Failed	Total	Agreement %
GSC	'most likely'	14	1	15	93.3%
GSC	'probable'	5	5	10	50.0%
Endeavor	'most likely'	61	2	63	96.8%
Endeavor	'probable'	38	23	61	62.3%
Combined	'most likely'	75	3	78	96.2%
Combined	'probable'	43	28	71	60.6%

5. Discussion

The CIT was developed in order to support future investigations into hydrothermal venting in seafloor spreading ridge environments, specifically those that benefit from the ability to find individual chimneys or make a more complete inventory of both active and inactive chimneys. Examples of these types of investigations include the recent discovery of off-axis hydrothermal venting at the 9° N segment of the EPR [34], an investigation the relationship between hydrothermal venting and eruptive fissures [35,36], and a study of the role of faults in hydrothermal vent field locations [3]. Previous investigations, where ≤1 m gridded AUV bathymetry data exist, currently rely on the time-consuming manual picking of hydrothermal vents to achieve these goals and, consequently, are uncommon in the literature (e.g., [7,33]). In the context of exploration efforts, AUV surveys are becoming commonplace for hydrothermal vent research, and yet, pinpointing locations for near-bottom work is crucial [4]. One of the advantages that the CIT provides for future research into hydrothermal systems is that the classification method is consistent and repeatable, meaning that differences in the observations between seafloor environments can be attributed to environmental variables rather than variations in the classification procedure by humans.

The two output sets of classified chimneys from the CIT, 'most likely chimneys' and 'probable chimneys', can be used in different ways to identify chimneys in areas of exploration. For example, utilizing the 'most likely chimneys' output from the CIT provides a way to quickly identify candidates for follow-up investigations (e.g., sampling or visual imaging) from exploratory AUV mapping because of the low commission error. The 'probable chimneys' group of the CIT output provides the ability to not only identify areas with hydrothermal chimneys present, but also to confidently identify those areas where there are no chimneys due to the low omission error.

The accurate classification of chimneys by the CNN is the most important aspect of the CIT, and we found that changes to the training dataset had the largest impact on the CNN model. The most drastic improvements to CNN performance were observed with the addition of more training classes, so that the CNN could learn to recognize and exclude features that are similar to but not hydrothermal chimneys. When initially constructed, the CNN was trained using two sets of examples: locations from the manual-pick datasets from each study area, and a set of randomly selected points from each dataset that were not chimneys. Visual inspection of the results of this version of the CIT showed many misclassified points located on the upper edges of fault scarps. In order to directly address this error, we implemented four classes of training data: flat/gently sloping areas, chimneys, the upper edges of fault scarps, and rounded local maxima that were at the edges of larger features (Figure 7). The flat, fault scarp, and rounded training datasets were produced using manual classification. We were unable to continue this method of adding training classes to improve performance, because there was no longer a consistent group of features with shared characteristics showing up in the output. This area of seafloor geosciences is still developing, and only a limited number of hydrothermal chimneys are available for training at this time. However, with additional training data, future AUV mapping surveys in more areas with past or present hydrothermal activity could benefit the performance of the CNN developed here.

The comparison of the CIT classification with the manual picking of hydrothermal chimneys show some mismatch between these datasets (Figures 11 and 12). For the GSC, there are only two chimneys identified by the CIT as 'most likely chimneys' that are not in the White and Lee [8] dataset, and several of the manual picks are also included in the 'probable chimneys' group. However, in the Endeavor test area, the CIT output sets and Clague et al.'s [7] manual picks show several examples of spatial disagreement (Figure 11) despite only two of the CIT outputs being caught by our manual classification in this study (Table 2). There are 51 Clague et al. [7] picks within this test area and 63 'most likely' CIT outputs, and 34 of them are shared. Some features that strongly resemble a chimney are included in the CIT output but not in the Clague et al. [7] manual picks, and a hydrothermal mound without a spire shape was included in the pre-existing dataset but not the CIT output (Figure 13). One potential reason for this is that Clague et al. [7] used other data to supplement the AUV bathymetry where available, such as ROV observations, to improve their hydrothermal chimney catalog. The CIT is designed to identify features that have the characteristics of a hydrothermal chimney from only the AUV bathymetry, which is the data available for most areas of the seafloor. Because the previous manual classifications use other observations and are interpretations made by expert human judgement, an imperfect match is expected. Additionally, we analyzed the CIT outputs by manually evaluating each output ourselves—instead of just comparing the output locations to the pre-existing datasets—as a basis for comparison to human picks from bathymetry alone, without the aid of a priori knowledge.
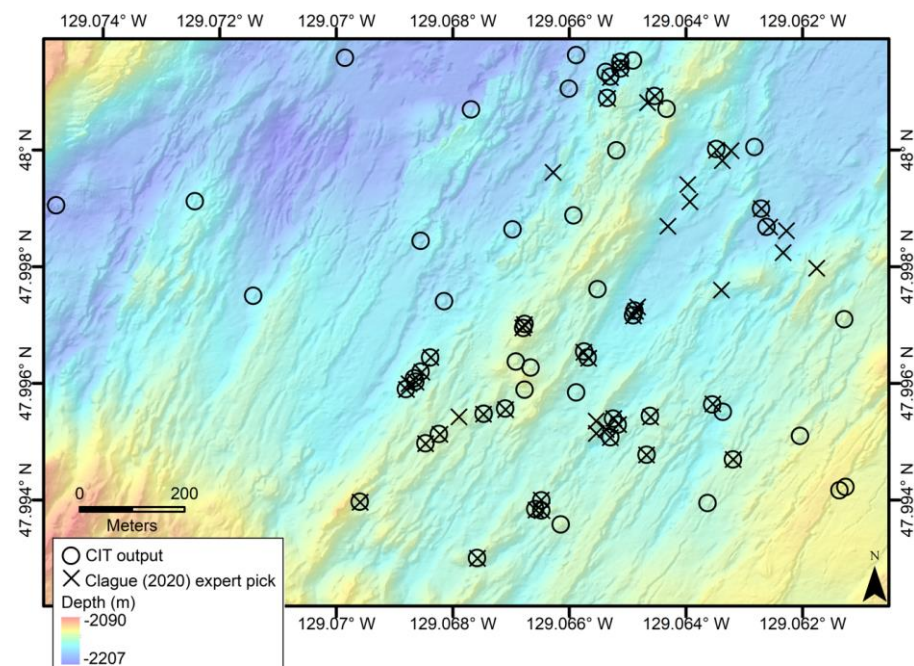
Figure 11. Shaded-relief bathymetry map showing the CIT 'most likely chimneys' output (open circles) for the Endeavor test area, along with the locations from the previous manual picks by Clague et al. [7] (black Xs). Although most of the chimney locations are the same in both datasets, there are multiple examples missed in one of the two datasets.
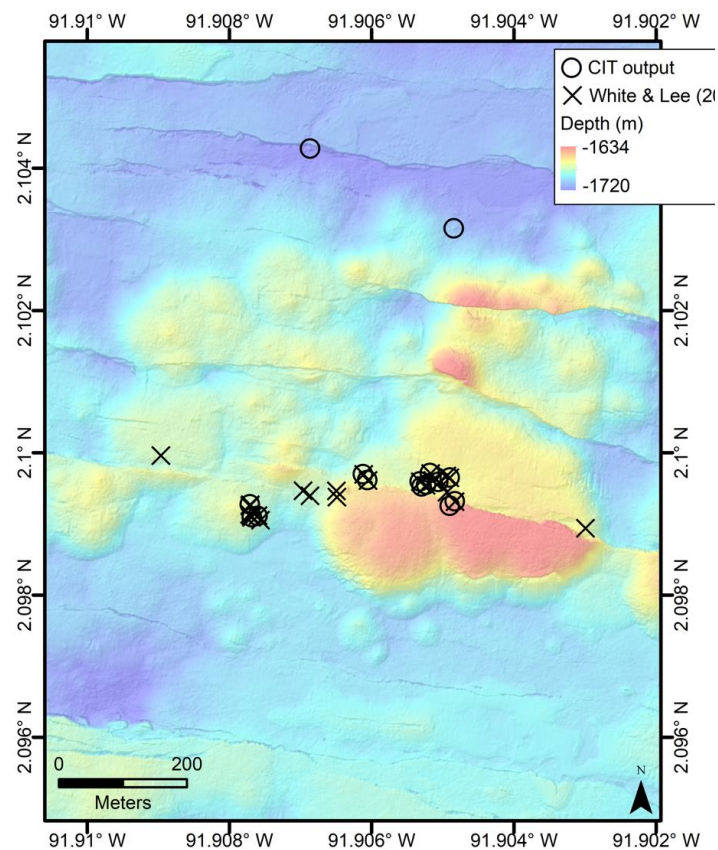
Figure 12. The CIT 'most likely chimneys' output for the GSC test area, along with the locations from the previous manual picks by White and Lee [8], on a basemap of shaded-relief 1 m gridded bathymetry. There is much less disagreement between the pre-existing manual picks and the CIT output, although the White and Lee [8] dataset contains many smaller chimneys not found by the CIT.
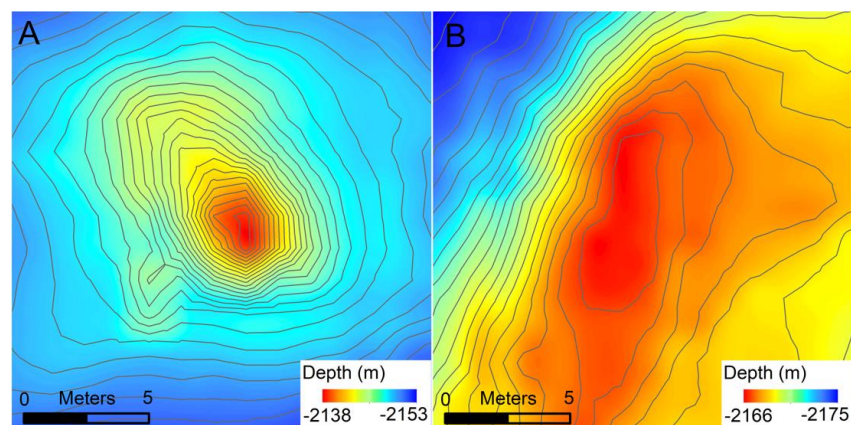
Figure 13. Bathymetric maps (0.5 m contour lines) showing a comparison of a feature identified by the CIT that is not included in the Clague et al. [7] manual-pick dataset (**A**) and an example of a feature from Clague et al. [7] that was not identified by the CIT (**B**). Both of these examples are considered accurate classifications by the CIT compared to the criteria used in this study; however, the previous manual picks were more inclusive of features that have atypical bathymetric expression.

The comparison of the CIT classifications and previous manual picks highlights the diversity of features, such as chimneys, that appear in meter-scale bathymetry of seafloor spreading ridges. Distinguishing hydrothermal chimneys in bathymetry is a difficult task not only for a Machine-Learning tool, but for humans as well. Clague et al. [7] discuss how

both lava pillars and mounds of pillow lava were included in their preliminary manual picks of chimneys on the Endeavor segment, and although they removed many, they expect that some still remain in their results. This point underscores the need for accuracy assessment inherent in all classifications from remote-sensing data [37]. The CIT performs well into the 80% accuracy range, but consideration of the potential for misclassification of features is important for future applications of the CIT. Areas of complex bathymetry, faulted grabens, calderas, and collapsed axial summit troughs will pose challenges for hydrothermal chimney detection from bathymetric mapping alone.

6. Conclusions

This CIT was created to find hydrothermal chimneys from meter-scale bathymetry alone by training a Machine-Learning classifier. The CIT generates a primary group of locations, labeled as 'most likely chimneys'; it has been shown to catch a large majority of known hydrothermal chimneys (81%), and to only include features that were included in manual picks by expert human interpreters (96%). The CIT also produces a secondary group of 'probable chimneys' that identifies all known chimneys in our study areas, except for two very small chimneys with footprints of only 1–2 pixels; however, these are identified at the cost of including many more classified chimneys than what we found with manual classification (61% agreement). Having the ability to utilize both of these output groups provides flexibility in the use cases of the CIT, for applications requiring the identification of targets that are 'most likely chimneys' or for a more inclusive cataloguing of every 'probable chimney'. Benefits to the CIT are the consistency of the classification and the efficiency of automated classification compared to human interpretation. The CIT will aid research efforts focused on the distribution of hydrothermal venting, comparisons between different ridge environments, and the connections to underlying ridge geology, by providing an efficient and repeatable tool to identify hydrothermal chimney locations from 1m resolution gridded bathymetry.

Author Contributions: Both authors, I.K. and S.W., contributed to every aspect of this study, including: conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The processed AUV bathymetric data used in this study were accessed at the IEDA Marine Geosciences data repository at marine-geo.org (accessed on 20 September 2020). The Python scripts created for this project are available at the GitHub repository https://github.com/IPKeohane/Chimney-Identification-Tool.git, accessed on 15 January 2022 Other data and files generated in this study can be made available upon request to the corresponding author I.K. (ikeohane@seoe.sc.edu).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Baker, E.T.; German, C.R.; Elderfield, H. Hydrothermal Plumes Over Spreading-Center Axes: Global Distributions and Geological Inferences. In *Seafloor Hydrothermal Systems: Physical, Chemical, Biological, and Geological Interactions (Volume 91)*; Humphris, S.E., Zierenberg, R.A., Mullineaux, L.S., Thomson, R.E., Eds.; American Geophysical Union Monograph: Washington DC, USA, 1995; pp. 47–71.
2. Delaney, J.R.; Robigou, V.; McDuff, R.E.; Tivey, M.K. Geology of a Vigorous Hydrothermal System on the Endeavour Segment, Juan de Fuca Ridge. *J. Geophys. Res.* **1992**, *97*, 19663. [CrossRef]

3.   Glickson, D.A.; Kelley, D.S.; Delaney, J.R. Geology and Hydrothermal Evolution of the Mothra Hydrothermal Field, Endeavour Segment, Juan de Fuca Ridge. *Geochem. Geophys. Geosystems* **2007**, *8*, Q06010. [CrossRef]

4.   German, C.R.; Yoerger, D.R.; Jakuba, M.; Shank, T.M.; Langmuir, C.H.; Nakamura, K. Hydrothermal Exploration with the Autonomous Benthic Explorer. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **2008**, *55*, 203–219. [CrossRef]

5.   Jamieson, J.W.; Clague, D.A.; Hannington, M.D. Hydrothermal Sulfide Accumulation along the Endeavour Segment, Juan de Fuca Ridge. *Earth Planet. Sci. Lett.* **2014**, *395*, 136–148. [CrossRef]

6.   Paduan, J.B.; Zierenberg, R.A.; Clague, D.A.; Spelz, R.M.; Caress, D.W.; Troni, G.; Thomas, H.; Glessner, J.; Lilley, M.D.; Lorenson, T.; et al. Discovery of Hydrothermal Vent Fields on Alarcón Rise and in Southern Pescadero Basin, Gulf of California. *Geochem. Geophys. Geosystems* **2018**, *19*, 4788–4819. [CrossRef]

7.   Clague, D.A.; Martin, J.F.; Paduan, J.B.; Butterfield, D.A.; Jamieson, J.W.; Le Saout, M.; Caress, D.W.; Thomas, H.; Holden, J.F.; Kelley, D.S. Hydrothermal Chimney Distribution on the Endeavour Segment, Juan de Fuca Ridge. *Geochem. Geophys. Geosystems* **2020**, *21*, e2020GC008917. [CrossRef]

8.   White, S.; Lee, A. Hydrothermal Chimney Distribution from AUV Sentry Bathymetry and Alvin at the Galapagos Spreading Center. In Proceedings of the Fall Meeting AGU, AGU, San Francisco, CA, USA, 17 December 2014; p. V31B-4741.

9.   Ferrini, V.L.; Tivey, M.K.; Carbotte, S.M.; Martinez, F.; Roman, C. Variable Morphologic Expression of Volcanic, Tectonic, and Hydrothermal Processes at Six Hydrothermal Vent Fields in the Lau Back-Arc Basin. *Geochem. Geophys. Geosystems* **2008**, *9*, Q07022. [CrossRef]

10.   Géron, A. *Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*; Tache, N., Ed.; O'Reilly Media: Sebastopol, CA, USA, 2017; ISBN 9781491962299.

11.   Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]

12.   Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a Convolutional Neural Network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; pp. 1–6.

13.   O'Shea, K.; Nash, R. An Introduction to Convolutional Neural Networks. *arXiv* **2015**, arXiv:1511.08458.

14.   Bilbrey, J.A.; Heindel, J.P.; Schram, M.; Bandyopadhyay, P.; Xantheas, S.S.; Choudhury, S. A Look inside the Black Box: Using Graph-Theoretical Descriptors to Interpret a Continuous-Filter Convolutional Neural Network (CF-CNN) Trained on the Global and Local Minimum Energy Structures of Neutral Water Clusters. *J. Chem. Phys.* **2020**, *153*, 024302. [CrossRef]

15.   Jung, Y.-J.; Han, S.-H.; Choi, H.-J. Explaining CNN and RNN Using Selective Layer-Wise Relevance Propagation. *IEEE Access* **2021**, *9*, 18670–18681. [CrossRef]

16.   Juliani, C. Automated Discrimination of Fault Scarps along an Arctic Mid-Ocean Ridge Using Neural Networks. *Comput. Geosci.* **2019**, *124*, 27–36. [CrossRef]

17.   Marsh, I.; Brown, C. Neural Network Classification of Multibeam Backscatter and Bathymetry Data from Stanton Bank (Area IV). *Appl. Acoust.* **2009**, *70*, 1269–1276. [CrossRef]

18.   McClinton, J.T.; White, S.M.; Sinton, J.M. Neuro-Fuzzy Classification of Submarine Lava Flow Morphology. *Photogramm. Eng. Remote Sens.* **2012**, *78*, 605–616. [CrossRef]

19.   Valentine, A.; Kalnins, L. An Introduction to Learning Algorithms and Potential Applications in Geomorphometry and Earth Surface Dynamics. *Earth Surf. Dyn.* **2016**, *4*, 445–460. [CrossRef]

20.   Valentine, A.P.; Kalnins, L.M.; Trampert, J. Discovery and Analysis of Topographic Features Using Learning Algorithms: A Seamount Case Study. *Geophys. Res. Lett.* **2013**, *40*, 3048–3054. [CrossRef]

21.   DeMets, C.; Gordon, R.G.; Argus, D.F. Geologically Current Plate Motions. *Geophys. J. Int.* **2010**, *181*, 1–80. [CrossRef]

22.   McClinton, J.T.; White, S.M. Emplacement of Submarine Lava Flow Fields: A Geomorphological Model from the Niños Eruption at the Galápagos Spreading Center. *Geochem. Geophys. Geosystems* **2015**, *16*, 899–911. [CrossRef]

23.   White, S.M.; Meyer, J.D.; Haymon, R.M.; Macdonald, K.C.; Baker, E.T.; Resing, J.A. High-Resolution Surveys along the Hot Spot-Affected Galápagos Spreading Center: 2. Influence of Magma Supply on Volcanic Morphology. *Geochem. Geophys. Geosystems* **2008**, *9*, Q09004. [CrossRef]

24.   Clague, D.A.; Dreyer, B.M.; Paduan, J.B.; Martin, J.F.; Caress, D.W.; Gill, J.B.; Kelley, D.S.; Thomas, H.; Portner, R.A.; Delaney, J.R.; et al. Eruptive and Tectonic History of the Endeavour Segment, Juan de Fuca Ridge, Based on AUV Mapping Data and Lava Flow Ages. *Geochem. Geophys. Geosystems* **2014**, *15*, 3364–3391. [CrossRef]

25.   Le Saout, M.; Clague, D.A.; Paduan, J.B. Evolution of Fine-Scale Segmentation at Intermediate-Spreading Rate Ridges. *Geochem. Geophys. Geosystems* **2019**, *20*, 3841–3860. [CrossRef]

26.   Ryan, W.B.F.; Carbotte, S.M.; Coplan, J.O.; O'Hara, S.; Melkonian, A.; Arko, R.; Weissel, R.A.; Ferrini, V.; Goodwillie, A.; Nitsche, F.; et al. Global Multi-Resolution Topography Synthesis. *Geochem. Geophys. Geosystems* **2009**, *10*, Q03014. [CrossRef]

27.   Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2014**, 580–587. [CrossRef]

28.   Bycroft, R.; Leon, J.X.; Schoeman, D. Comparing Random Forests and Convoluted Neural Networks for Mapping Ghost Crab Burrows Using Imagery from an Unmanned Aerial Vehicle. *Estuar. Coast. Shelf Sci.* **2019**, *224*, 84–93. [CrossRef]

29.   Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv* **2019**, arXiv:1912.01703.

30. Walbridge, S.; Slocum, N.; Pobuda, M.; Wright, D.J. Unified Geomorphological Analysis Workflows with Benthic Terrain Modeler. *Geosciences* **2018**, *8*, 94. [CrossRef]

31. Simard, P.Y.; Steinkraus, D.; Platt, J.C. Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. In Proceedings of the Seventh International Conference on Document Analysis and Recognition, Edinburgh, UK, 6 August 2003; Volume 1, pp. 958–963.

32. Bottou, L. Stochastic Gradient Learning in Neural Networks. *Proc. Neuro-Nımes* **1991**, *91*, 12.

33. Hecht-Nielsen, R. Theory of the Backpropagation Neural Network. In *Neural Networks for Perception*; Wechsler, H., Ed.; Academic Press: Cambridge, MA, USA, 1992; pp. 65–93. ISBN 978-0-12-741252-8.

34. Wu, J.; Parnell-Turner, R.; Fornari, D.J.; Kurras, G.; Berrios-Rivera, N.; Barreyre, T.; McDermott, J.M. Extent and Volume of Lava Flows Erupted at 9°50′ N, East Pacific Rise in 2005–2006 from Autonomous Underwater Vehicle Surveys. *Geochem. Geophys. Geosyst.* **2022**, *23*, e2021GC010213. [CrossRef]

35. Hey, R.; Baker, E.; Bohnenstiehl, D.; Massoth, G.; Kleinrock, M.; Martinez, F.; Naar, D.; Pardee, D.; Lupton, J.; Feely, R.; et al. Tectonic/Volcanic Segmentation and Controls on Hydrothermal Venting along Earth's Fastest Seafloor Spreading System, EPR 27°–32° S. *Geochem. Geophys. Geosystems* **2004**, *5*, Q12007. [CrossRef]

36. Wright, D.J.; Haymon, R.M.; Fornari, D.J. Crustal Fissuring and Its Relationship to Magmatic and Hydrothermal Processes on the East Pacific Rise Crest (9°12′ to 54′ N). *J. Geophys. Res. Solid Earth* **1995**, *100*, 6097–6120. [CrossRef]

37. Congalton, R.G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sens. Environ.* **1991**, *37*, 35–46. [CrossRef]