Machine Learning for Automated Seabed Mapping

Umberto Di Laudo[1,2,*], Silvia Ceramicola[2] and Luca Manzoni[1,2]

[1]Dipartimento di Matematica, Informatica e Geoscienze, Università degli Studi di Trieste, Via Alfonso Valerio 12/1, 34127 Trieste, Italy
[2]Istituto Nazionale di Oceanografia e Geofisica Sperimentale, Borgo Grotta Gigante 42/c, 34010 Sgonico, Italy

Abstract

Interpreting morphological features of the seabed is a labor-intensive task for marine geologists especially when it concerns extensive portions of seabed. By applying Machine Learning (ML) techniques from the field of computer vision, it is possible to significantly streamline this process, speeding it up considerably. In this paper we present a model capable of automatically categorizing seabed features, identifying different morphological elements, such as submarine canyons, escarpments, canyon headwalls and mass movements. This model will serve as the basis for new tools to assist geologists as well as stakeholders dealing with management of coastal or offshore areas in their work, providing them with an efficient support for seabed analysis and characterization.

Keywords

Deep Learning, Seabed Mapping, Image Segmentation

1. Introduction

One important task for marine geologist is the identification of the morphological characteristics of the seabed. Underwater morphological features, such as canyons, escarpments, are components of the seabed environment. These elements are typically shaped by geological processes, including erosion, sedimentation, and tectonic activity, over extended periods. Detecting their occurrence and characteristics play a crucial role in assessing marine hazards or when placing communication cables at seabed [1, 2]. However, detection and mapping of these underwater morphological elements require specialized domain expertise, such as marine geologist. Despite this process being very time-consuming for scientists, nowadays there is still no automated method to detect and classify the various elements of the underwater environment.

Hence, any support to help the automatic identification of these features via machine learning (ML) would produce a significant benefit for marine geologists, making the entire process smoother and less time-consuming. The overarching aim of this project is to develop a model capable of autonomously classifying seabed features, and thus assisting geologists in their analysis. This model will be designed to identify various morphological elements present in seabed data, like submarine canyons, escarpments, mass movements. To do so, the task of seabed interpretation can be considered as a special case of image segmentation [3], where images are replaced with a map of seabed features and the labeling associated to each pixel corresponds to the morphological feature of the seabed in the specific latitude and longitude. Hence, standard image segmentation techniques can be employed and adapted for this task. In this paper we present the first results of using a *U-Net* [4] architecture for seabed classification trained using the data obtained via the *MaGIC* project (*Marine Geohazards along the Italian Coasts*) [5, 6], which produced maps of the Italian coasts of Central and South Italy, Sicily, Sardinia, and Liguria in a five-year time frame starting from 2007.

The paper is structured as follows: in Section 2 the current state if the art in image segmentation and the main groups of existing techniques are presented. In Section 3 the available data are presented and the specific task to be solved is further detailed. The architecture of the network and the training process are detailed in Section 4. The results are then presented in Section 5. Finally, in Section 6 we present the planned research directions.

2. Image Segmentation

Image segmentation is a computer vision technique that involve partitioning a digital image into multiple segments or regions, separating meaningful objects or structures within an image from the background or other objects, i.e., it can be considered a classification problem at pixel-level. The segmentation process divides images into different regions based on certain characteristics such as color, intensity, texture, or other features. In particular, with the advent of powerful ML techniques, the features are learned by the data instead of being handcrafted by experts.

Ital-IA 2024: 4th National Conference on Artificial Intelligence, organized by CINI, May 29-30, 2024, Naples, Italy
*Corresponding author.
†These authors contributed equally.
udilaudo@ogs.it (U. Di Laudo); sceramicola@ogs.it (S. Ceramicola); lmanzoni@units.it (L. Manzoni)
0000-0002-1318-1272 (S. Ceramicola); 0000-0001-6312-7728 (L. Manzoni)

According to [3], there exists three groups of image segmentation:

- Semantic segmentation: Semantic segmentation involves classifying each pixel in an image into a specific category or class.
- Instance segmentation: it extends semantic segmentation by not only classifying each pixel into categories but also distinguishing between different object instances of the same category.
- Panoptic segmentation: it aims to unify semantic segmentation and instance segmentation into a single framework. It divides an image into semantically meaningful regions and assigns a unique label to each region, regardless of whether it corresponds to an object instance or a background category.

The main traditional image segmentation techniques include several methods like thresholding, histograms, watersheds, region-growing and clustering based segmentation [7].

As stated above, with the advent of deep learning, new techniques for image segmentation were developed [7], with the main distinction usually being the architecture of the neural network used. One of the first architecture used for semantic segmentation was a *Fully Convolutional Network* (FCN) proposed by Long *et al.* in 2015 [8]. That network includes only convolutional layers and is able to take an image of a certain size as input and returns a segmentation map of the same dimensions. Other deep learning-based models for segmentation was proposed following the encoder-decoder architecture: Badrinarayanan *et al.* proposed the *SegNet* [9].

Inspired by the FCN and the encoder-decoder architecture, *V-Net* [10] and *U-Net* [4] were proposed, initially mainly for medical and biomedical purposes. Over the years, several modification of the *U-Net* architecture were performed to adapt it to different kind of images. For example Zhou *et al* [11] proposed a nested *U-Net* architecture. Furthermore Cicek [12] build a *U-Net* architecture for $3D$ images. Nowadays *U-Net* are also used in other fields, e.g., road segmentation [13], face detection [14], and autonomous driving [15].

3. Seabed Data

In this section we present the main characteristics of the data used for training and testing the proposed model.

3.1. Input Data

The data are provided by the *Italian MaGIC project* are GIS data specifying the *depth* of the shallow coastal regions in Italy [16, 17, 18]. Starting from depth data it is possible to derive via standard GIS tools two additional features that are considered useful by domain experts: the *slope* and the *profile curvature* of the seabed, that are respectively the first and the second derivative of the depth. The corresponding data can then be interpreted as $2D$ fields with three features/channels (i.e., it can be directly interpreted/visualized as an image), each one describing a specific feature of the same seabed area. The data are organized in a first image of about 2800×2400 pixels, while the second of 3000×3800 pixels. Each of then is cut into smaller square windows of length 100×100 pixels for the training process, for a total of more than 2100 squares.

3.2. Ground Truth

All the data obtained by the MaGIC project also received a human interpretations of the seabed structure i.e., a *labels map* of the same dimensions of the input 2D fields indicating the positions of all the elements present in that region. In particular, the labeling is done by drawing lines of different types over the depth maps. The different types of lines corresponds to different classes which are 97 in the original data. Due to the large number of classes and the fact that many of them were only represented by a small number of samples, a first pre-processing step was done by reducing the number of classes to 15. These 15 classes corresponds to grouping of the original 97 classes with the partitioning done according to the morphological and geological similarity of the seabed features and with the help of marine geologists. Hence, the output of the model has to be a $2D$ field in which each coordinate (each pixel) belongs to one of the 16 classes (15 related to the different morphological elements plus 1 for the background).

4. Architecture and Training

In this work we perform *semantic segmentation* using a *U-Net* architecture [19]. The network architecture is comprised of three main elements 1:

- **A contracting path** that reduce the spatial dimensions of the input image;
- **An expansive path** that increase the spatial dimensions;
- **Skip connection** between corresponding layers in the contracting and expansive part.

The specific architecture used in this study is presented in Figure 1, with the parameters of the different layers presented in Table 1. The contracting path has the typical structure of a Convolutional Neural Network (CNN), which in our case it consist in five layers. The first has three input channels (corresponding to the three input

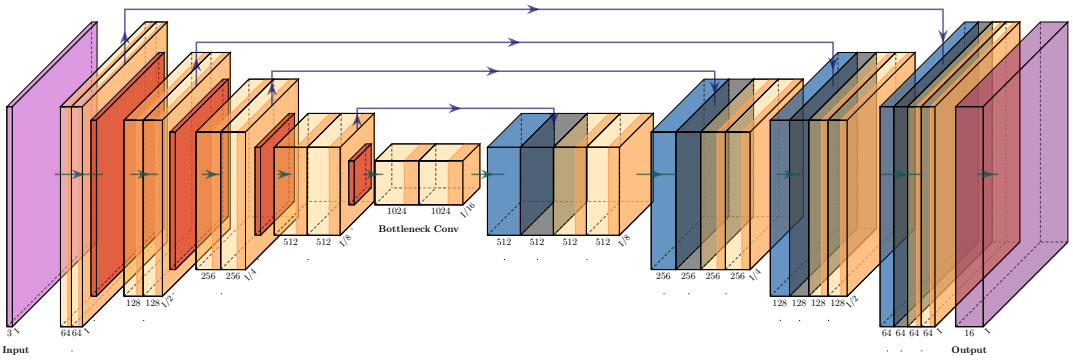

Figure 1: U-Net architecture used in this work. The yellow and blue boxes represent convolutional and deconvolutional layers, respectively. The *ReLU* functions are represented by the orange part of the boxes. The red squares corresponds to *max pooling* layers.

features) and performs two convolutions followed by batch normalization and *ReLU* activation function. As stated above, the input data have a size of 100×100 pixels. The remaining layers of this part has the same structure with the addition of a *max pooling* operation at the beginning.

Table 1
Convolutional and deconvolutional layers structure

Type	Kernel size	Padding	Stride
Convolution	3	1	1
Deconvolution	2	0	2

Then there is the expansive path composed by five layers in which deconvolution are performed in order to upscale the image. The last layer has 16 output channels like the total number of classes of seabed's elements. The output values are logit, which, if necessary, can then be transformed in a probability distribution over the different classes via softmax.

4.1. Training

As stated before, the network is trained with 100×100 pixels images, cut from the total *features map*. Hence, this network take a 100×100 3-channels image as input and returns a 100×100 16-channels tensor. In this framework for each pixel we have 16 number, each one related to a specific class. The class related to the channel with the bigger number is associated to that pixel.

The loss between the output of the *U-Net* and the transformed *labels map* is computed by using a traditional *Cross-Entropy loss* in addition to the *Dice loss* [20]. The last one is a particular loss usually used in segmentation problem that is related to the intersection between the prediction and the target images.

We partitioned the dataset into an $80\% - 20\%$ split for training and testing, respectively and used a batch size of 128. We employed the Adam optimizer with a learning rate set to 0.001 and the training continued for 60 epochs.

Notice that the actual classification is *not* done directly by assigning the most probable class, but by assigning a threshold for the logit value for all the 15 classes corresponding to actual morphological features of the seabed. If one of them is above the threshold then the assigned classes is the one corresponding to it, even if the background (i.e., no feature present) has a higher logit value and would have been the most probable. This is done because of the strong imbalance of the dataset (the "no feature" class is present for more than 98% of the pixels) and, thus, the model would prefer selecting that class.

5. Results

In this section we present an initial analysis of the results obtained.

In Figure 2 two *labels maps* are presented. They correspond to the first region in which the total dataset was split: in the left one there are the human interpretations of the underwater environment provided by the *MaGIC project* that map every relevant element in the area; the right one is the map reconstructed by the network after the training. The the white squares denote the test set that represent the 20% of the total part of the region.

In Table 2 a comparison between the frequencies of the labels of the reconstructed map and the human interpretations (the ground truth) is given. The frequencies excludes the background pixels, i.e., where no feature is present.

Figure 2: In the left figure to the left there is the human interpretation of the seabed of one the two regions in which the dataset is split. To the right there is the reconstructed map; the white squares represent the test set, which is not used during the training.

Table 2
Comparison between the frequencies of the labels of the reconstructed map and the ground truth without considering the background. Only the test areas are considered.

Label	Ground Truth freq.	Reconstructed map freq.
1	11.93%	9.46%
2	1.18%	1.19%
3	22.56%	33.93%
4	17.33%	17.30%
5	0.00%	0.18%
6	0.00%	0.00%
7	5.82%	1.10%
8	0.40%	0.00%
9	14.50%	13.64%
10	4.86%	0.35%
11	0.00%	0.00%
12	15.39%	21.74%
13	0.14%	0.00%
14	0.00%	0.00%
15	5.89%	1.10%

Table 3
Comparison between the frequencies of the labels of the reconstructed map with no threshold and threshold value equal to -1.5 and the ground truth. Only the test areas are considered.

Label	Ground Truth	Output map	Threshold of -1.5
0	98.05%	99.75%	98.37%
1	0.23%	0.02%	0.15%
2	0.02%	0.00%	0.02%
3	0.44%	0.08%	0.50%
4	0.34%	0.04%	0.20%
5	0.00%	0.00%	0.00%
6	0.00%	0.00%	0.00%
7	0.11%	0.00%	0.07%
8	0.01%	0.00%	0.00%
9	0.28%	0.03%	0.31%
10	0.09%	0.00%	0.03%
11	0.00%	0.00%	0.00%
12	0.30%	0.05%	0.30%
13	0.00%	0.00%	0.00%
14	0.00%	0.00%	0.00%
15	0.11%	0.00%	0.05%

Recall that due to the predominance of one of "no features" class over the others, the model tends to favour this class, as typical for highly imbalanced datasets. To show that the use of threshold actually helps in improving the classification we compared the class frequencies with and without a threshold. In Table 3 frequencies of the classes of the ground truth are compared with those of the output map with no threshold and with $threshold = -1.5$ considering now also the background (label 0). Notice that the frequency of the pixels of the background decrease with adding the threshold and also the other frequencies are more similar to those of the ground truth.

The model trained with the data representing the region shown in Figure 2, that are only a part of the entire dataset, is used on another region to test the model. The results are shown in Figure 3: the image to the left is ground truth while the image to right is the reconstructed map with no threshold used. We can easily notice that a predominance of the background is present. In the Figure 4 results with different values of threshold acting on the logit of the *non-zero* classes are present. The reduction of the threshold value correspond to an increase in the presence of pixels belonging to non-zero labels. Clearly, the correct choice of a suitable threshold

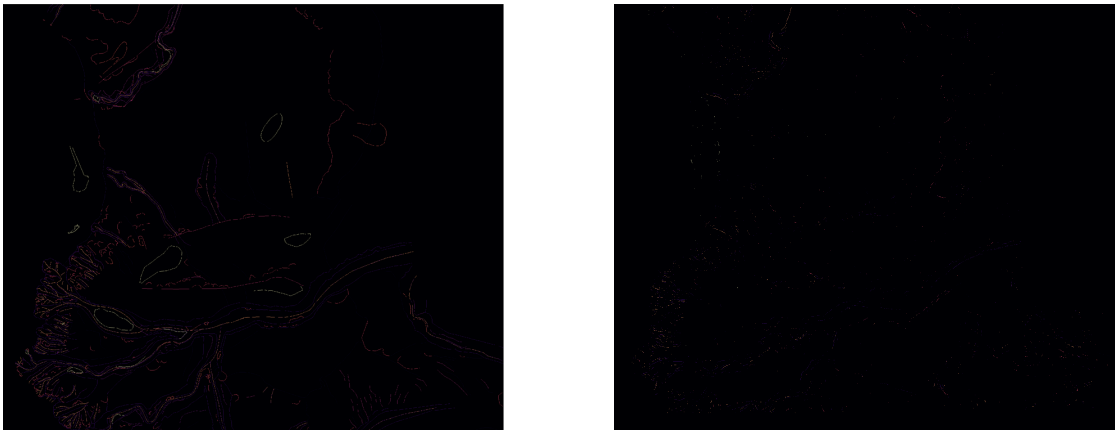

Figure 3: Comparison between the ground truth and the reconstructed map with no threshold. We can notice that there is a predominance of the zero class (background)

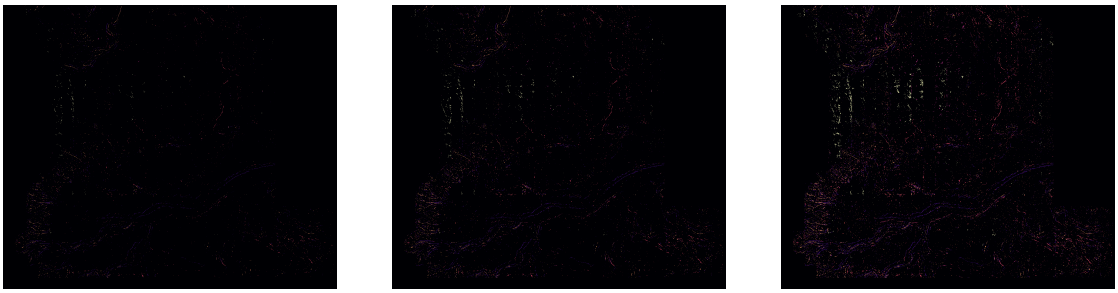

Figure 4: Comparison between the reconstructed maps with different values of the threshold. The value of the thresholds used are, from left to right equal to $1, -1, -3$. Pixel density increases as threshold value decreases

is essential for obtaining a good labeling of the seabed.

One important aspect to notice is that, for the proposed results, the actual loss is not a good indicator of the usefulness of the results. In fact, it represents only an proxy of the real usefulness of the proposed labeling, since labeling by experts is itself a subjective and noisy act. Thus, obtaining a feature that is shifted by a small amount might be immaterial (the original lines were themselves drawn by hand), like getting a non-continuous line (the expert interpretation would be that there should be a "connection" between two features). Hence, most of the evaluation is still qualitative and based on discussion with experts, that are particularly interested in the ability of the model to produce the correct "general shape" of the features, more than some specific details.

6. Conclusions and Future Work

The aim of this work was to construct a model that can identify and recognize the relevant morphological element of the seabed in an automatic way. The current results shows that, by using a U-net it is possible to produce good results from a qualitative point of view, as long as thresholding is used in the output of the network. Future goals are to improve the performance of the existing model in order to obtain more precise results. In particular, a new quantitative measure to encode the expert knowledge should be devised in order to speed-up the evaluation of the model (and, possibly, as a loss function for the training). Additionally, a future project will be to construct a model capable not only of identifying and mapping morphological elements of the seabed but also detecting potentially dangerous zones within it, i.e., *geohazards*.

References

[1] F. Chiocci, D. Ridente, Regional-scale seafloor mapping and geohazard assessment. The experience from the Italian Project MaGIC (Marine Geohazards along the Italian Coasts), Marine Geophysical

Research, 2011. doi:https://doi.org/10.1007/s11001-011-9120-6.

[2] F. L. Chiocci, A. Cattaneo, R. Urgeles, Seafloor mapping for geohazard assessment: state of the art, Marine Geophysical Research, 2011. doi:https://doi.org/10.1007/s11001-011-9139-8.

[3] A. Kirillov, K. He, R. Girshick, C. Rother, P. Dollar, Panoptic segmentation, volume 2019-June, 2019. doi:10.1109/CVPR.2019.00963.

[4] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, volume 9351, 2015. doi:10.1007/978-3-319-24574-4_28.

[5] Progetto MaGIC - Marine Geohazards along the Italian Coasts, 2007-2012. URL: https://www.protezionecivile.gov.it/en/approfondimento/progetto-magic-marine-geohazards-along-italian-coasts-0/, accessed on 2024-04-15.

[6] S. Ceramicola, D. Praeg, M. Coste, E. Forlin, A. Cova, E. Colizza, S. Critelli, Submarine Mass-Movements Along the Slopes of the Active Ionian Continental Margins and Their Consequences for Marine Geohazards (Mediterranean Sea), Springer International Publishing, Cham, 2014, pp. 295–306. doi:10.1007/978-3-319-00972-8_26.

[7] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, D. Terzopoulos, Image segmentation using deep learning: A survey, IEEE Transactions on Pattern Analysis and Machine Intelligence 44 (2022) 3523–3542. doi:10.1109/TPAMI.2021.3059968.

[8] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, volume 07-12-June-2015, 2015. doi:10.1109/CVPR.2015.7298965.

[9] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2017) 2481–2495. doi:10.1109/TPAMI.2016.2644615.

[10] F. Milletari, N. Navab, S. A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, 2016. doi:10.1109/3DV.2016.79.

[11] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, volume 11045 LNCS, 2018. doi:10.1007/978-3-030-00889-5_1.

[12] Özgün Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger, 3d u-net: Learning dense volumetric segmentation from sparse annotation, volume 9901 LNCS, 2016. doi:10.1007/978-3-319-46723-8_49.

[13] Z. Zhang, Q. Liu, Y. Wang, Road extraction by deep residual u-net, IEEE Geoscience and Remote Sensing Letters 15 (2018) 749–753. doi:10.1109/LGRS.2018.2802944.

[14] K. Luu, C. Zhu, C. Bhagavatula, T. H. N. Le, M. Savvides, A Deep learning approach to joint face detection and segmentation, 2016. doi:10.1007/978-3-319-25958-1_1.

[15] Çağrı Kaymak, A. Uçar, A brief survey and an application of semantic image segmentation for autonomous driving, volume 136, 2019. doi:10.1007/978-3-030-11479-4_9.

[16] S. Ceramicola, F. Fanucci, C. Corselli, E. Colizza, D. Morelli, A. Cova, A. Savini, D. Praeg, M. Zecchin, A. Caburlotto, O. Candoni, D. Civile, M. Coste, D. Cotterle, S. Critelli, A. Cuppari, M. Deponte, R. Dominici, E. Forlin, E. Gordini, C. Tessarolo, F. Marchese, F. Muto, S. Palamara, R. Riccardo, L. Facchin, R. Romeo, Calabria Ionica (Tavola 8, Fogli 35-39), in: C. FL, B. F, C. S, G. F, O. P. (Eds) (Eds.), Atlante dei lineamenti di pericolosità geologica dei mari italiani - Risultati del progetto MaGIC, CNR Edizioni, 2021, pp. 174–195.

[17] S. Ceramicola, M. R. Senatore, A. Cova, A. Meo, M. Zecchin, D. Praeg, C. Diego, C. Salvatore, A. Caburlotto, D. Civile, M. Coste, R. Dominici, E. Forlin, F. Muto, A. Bosman, C. Francesco Latino, E. Lai, D. Casalbore, E. Morelli, O. Candoni, E. Gordini, D. Michele, R. Riccardo, L. Facchin, R. Romeo, Golfo di Taranto (Tavola 9, Fogli 40-46), in: C. FL, B. F, C. S, G. F, O. P. (compilers) (Eds.), Atlante dei lineamenti di pericolosità geologica dei mari italiani - Risultati del progetto MaGIC, CNR Edizioni, 2021, pp. 196–225.

[18] C. F.L., F. Budillon, S. Ceramicola, F. Gamberi, P. Orrù (Eds.), Atlante dei lineamenti di pericolosità geologica dei mari italiani-Risultati del progetto MaGIC, CNR Edition, 2021.

[19] A. Milesi, Pytorch-UNet: Pytorch implementation of the u-net for image semantic segmentation with high quality images, 2024. URL: https://github.com/milesial/Pytorch-UNet.

[20] R. Zhao, B. Qian, X. Zhang, Y. Li, R. Wei, Y. Liu, Y. Pan, Rethinking dice loss for medical image segmentation, volume 2020-November, 2020. doi:10.1109/ICDM50108.2020.00094.