D4SC: Deep Supervised Semantic Segmentation for Seabed Characterisation in Low-Label Regime

Yoann Arhant[1,2*], Olga Lopera Tellez[1], Xavier Neyt[1], and Aleksandra Pižurica[2]

[1]Dept. of Communications, Information, Systems and Sensors, Royal Military Academy, Brussels

[2]Dept. of Telecommunications and Information Processing, Ghent University, Ghent

Abstract

Seabed characterisation consists in the study of the physical and biological properties of the bottom of the oceans. It is effectively achieved with sonar, a remote sensing method that captures acoustic backscatter of the seabed. Classical Machine Learning (ML) and Deep Learning (DL) research have failed to successfully address the automatic mapping of the seabed from noisy sonar data. This work introduces the Deep Supervised Semantic Segmentation model for Seabed Characterisation (D4SC), a novel U-Net-like model tailored to such data and low-label regime, and proposes a new end-to-end processing pipeline for seabed semantic segmentation. That dual contribution achieves state-of-the-art results on a high resolution Synthetic Aperture Sonar (SAS) survey dataset.

Keywords : Seabed, Semantic Segmentation, Synthetic Aperture Sonar, Deep Learning

1 Introduction

Seabed characterisation consists in the study of the physical and biological properties of submerged grounds. Its forms could range from divers and Remotely Operated Vehicles (ROVs) gathering visual intelligence on sediments, possibly taking optical pictures, to seabed sampling. Sonar, which succeeds to capture acoustic backscatter for wide areas, was proven particularly useful to detect boundaries between rather homogeneous seabeds[1] and for Automatic Target Recognition (ATR) within Mine Countermeasures (MCM) applications in shallow waters. Despite the remarkable detection accuracies of [2], that method also reported severe performance drops in particular seabeds, since sand ripples, rocky terrains and fields of seagrass cast shadows that might hide targets. Thus, seabed characterisation is paramount to assess the confidence of ATRs, but also for MCM mission planning. Our approach extends the work of [3], which addressed seabed characterisation as a pixel-wise semantic segmentation task from high resolution sonar data, to yield state-of-the-art results over real-world MCM operation datasets.

*This research was funded by grant DAP-21/11 from the Belgian Royal Higher Institute for Defence.

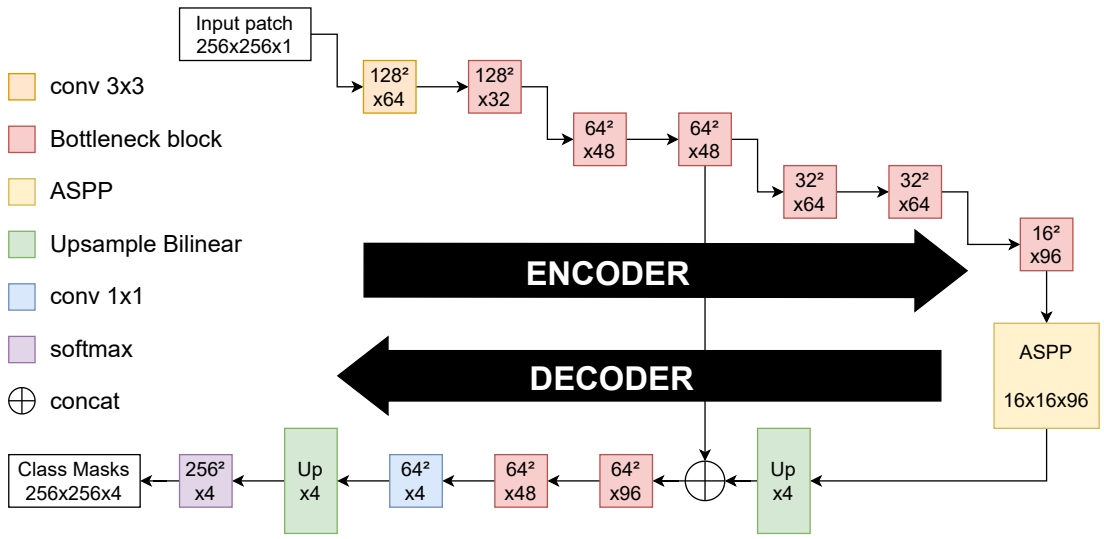


Figure 1: A visualisation of D4SC's architecture from its feature maps size after the different layers and operartions.

Deep Learning (DL) culminates with the application of Transformer models, reaching billions of parameters and capable of handling the training on datasets comprised of billions of images[4]. This is known as the high-data and high-label regime leading to models being highly robust to noise and outliers. Nevertheless, training on smaller datasets comprised of hundreds of images requires other learning schemes. Thus, the seabed characterisation literature investigated deep AutoEncoder (AE) structures, as their shrinkage to an embedded space is a regulariser in itself. For instance, [5], [6] and [7] trained a plain AE, a ladder network and a U-Net based model. Originally, the U-Net model[8] employed a symmetric Convolutional Neural Network (CNN) AE with skip connections, effectively guiding the reconstruction within the decoder part with lower level features from the encoder. Besides, [7] performed unsupervised pixel-wise segmentation with superpixels and transfer learning, striving to reuse the knowledge learnt from natural images distribution.

State-of-the-art methods employed notwithstanding, none of the aforementioned researches have succeeded in addressing wide-range automatic mapping of the seabed from low-label regime. Additionally, the input sonar data is hard to exploit due to the complex propagation of sound underwater and the different sources of acquisition errors and noise. The main one is the speckle noise, as a result of coherent imaging. To address both issues, we propose a novel U-Net-like model altogether with new data augmentation schemes. They blend in an end-to-end learning pipeline which achieves state-of-the-art results. It is introduced in Section 2. Then, Section 3 describes the SAS dataset. Section 4 presents D4SC's results over such data. Finally, section 5 concludes the paper.

2 Method

As the annotation procedure necessitates the support of experts or even in-field analyses, seabed characterisation falls in the low-label regime. To guide the learning phase, the literature is prone to

using either unsupervised, semi-supervised or self-supervised strategies. Other approaches strive to make the most of the knowledge of expert oracles and employ an iterative human-in-the-loop strategy called active learning. When addressing not only low-label but also low-data regimes, successful DL researches employ domain adaptation with transfer learning strategies. However, as acoustic tiles are rather homogeneous yet noisier than their visual counterpart, DL seabed characterisation will hardly benefit from this knowledge, while suffering from the massive amount of intermediate features to combine.

2.1 Architecture design

The overall design of our approach is meant to shred trainable parameters to avoid overfitting with low-label regime as it addresses seabed segmentation with limited classes and data. The proposed architecture, which is depicted in Fig. 1, employs a U-Net[8] encoder, but substitutes the last embedded space convolutional layer with an Atrous Spatial Pyramid Pooling (ASPP) block, feeding the features to a reduced decoder, as in Deeplab[9]. Additionally, we replace all remaining convolution layers, except the first one, by linear bottleneck blocks[10], that were proven state-of-the-art on embedded systems with limited cache memory. Besides, [11] showed them to be cutting-edge with different sizes and applications, which is all the more important to stabilize the training in low-label regime.

2.2 Data augmentation schemes

The deeper neural networks get, the more data they need to be fed with to achieve good results. One possible way to extend datasets is to perform thorough augmentations. While standard computer vision schemes for data augmentation consist in affine and color-space transforms, we extend them with a particular focus on the physical meaning of sonar data, depicted in Fig. 2. However employing them as is improved the classification results in computer vision challenges, seabed semantic

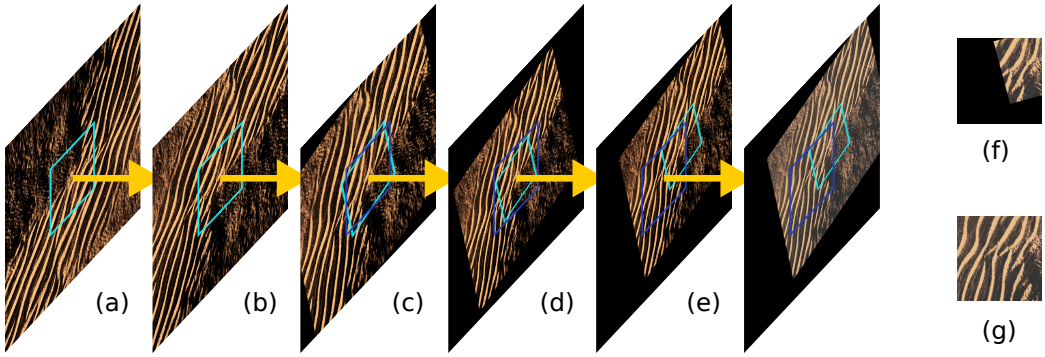


Figure 2: Diagram illustrating the detailed random data augmentation pipeline : (a) Flips, (b) Rotation, (c) Scaling, (d) Translation, (e) Intensity and Contrast Jittering. Following classical data augmentation schemes from computer vision, the resulting cropped patch represented by the dark blue square would only retain pixel information from the teal square, producing the (f) 256×256 pixels patch, whereas our method keep the cropped patch filled with natural input information (g).

segmentation would suffer from them as it would introduce non-physically based textures, non-noisy flat bottoms or impossible shadows in the input distribution. Therefore, we define pixel coordinates for patches and perform on-the-fly augmentations over them following standard computer vision schemes, but on bigger input patches, to ensure the final crop only spans natural input information – Fig. 2(g). Additionally, as shadows are only cast in range, it prevents us from applying anything but small random rotations to keep real-looking sonar images.

3 Description of the data

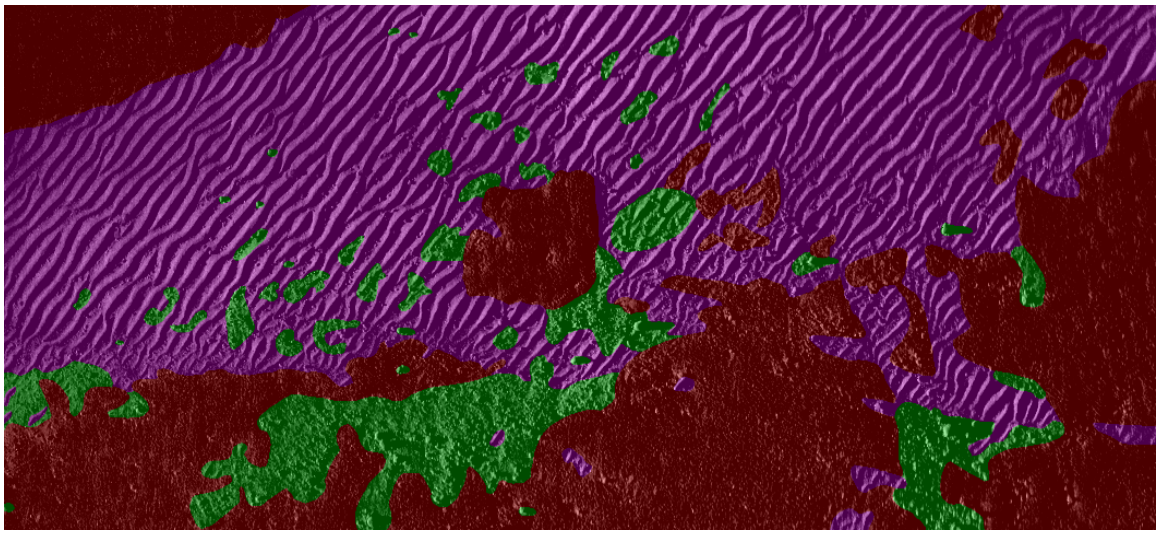
3.1 Description of the dataset

The Centre for Maritime Experimentation and Research (CMRE), conducted multiple surveys at sea over the past decade, with the MUSCLE Autonomous Underwater Vehicle (AUV) high-frequency SAS at an acquisition centre frequency of 300 kHz and gathered a large amount of data at a ground resolution up to 1.5 cm. One of those campaign was conducted in Latvia and mainly captured smooth flat bottom, sand ripples and rock outcrops altogether with underwater targets and confusers laid on the seabed within a MCM exercises scope. This dataset was best described in [2] and [3]. Numerous research publications addressed similar data, such as [12]. We follow its preprocessing pipeline performing median normalization in range and then azimuth over each image.

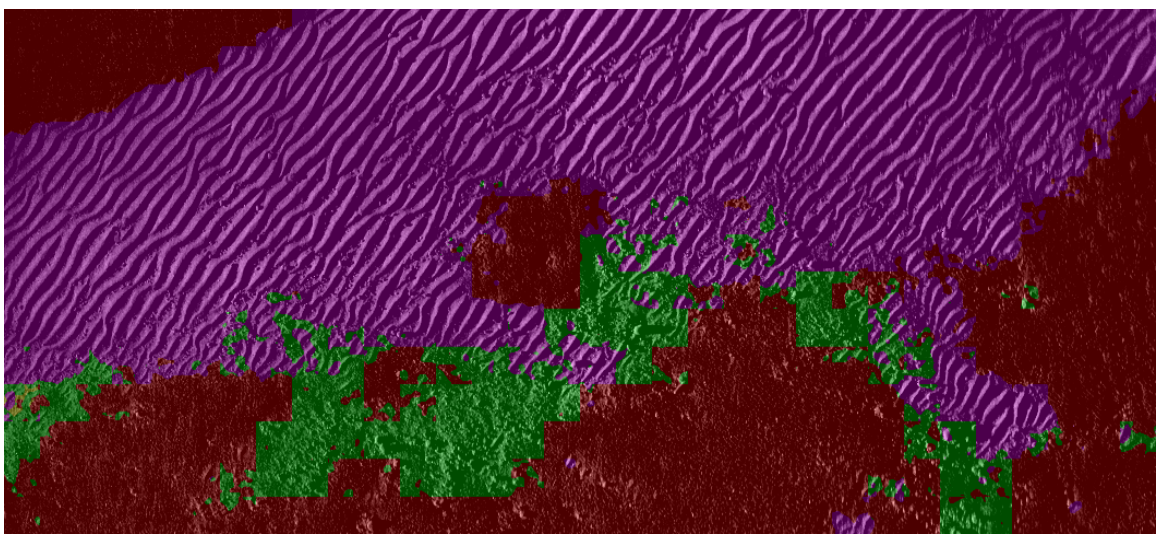
3.2 Manual Annotation

As the seabed is rather homogeneous within large areas for its characterisation with generic classes and only scarcely inhomogeneous, the polygon annotation strategy was selected to perform an efficient ground thruthing of seventeen randomly selected images of size 4400×2000 pixels. As they were spanning all natural and representative textures of the different classes but also multiple examples of particular seabed configurations from real data, such as a mega-ripple being high enough to cast shadows over a wide range, it was enough to start addressing seabed characterisation. Additionally, each manually drawn polygon was labelled with four disjoint classes, namely Flat Bottom, Rocks, Small Sand Ripples and Large Sand Ripples. As there is no vegetation in the whole survey, those classes embody the standard seabed classification scheme for MCM operations[3].

Although polygons were drawn to preserve boundaries between classes, as stated by [3], the annotation task is hard and sometimes the resulting borders are rather arbitrary. In addition, the tedious manual polygonal labelling task prevents annotators to dive deep into the image and to recover sediments surrounded by rock outcrops or the scarce boulders in the dead of large sandy areas. We restricted ourselves to patterns larger than 15 cm, being smaller than a MCM target. Besides, for complex seabeds describing the mixed aspect of an area, for instance a rock outcrop slowly emerging or being buried from smooth flat sand, the labelling procedure needs to bias one of the two resulting classes. Finally, it constitutes a good baseline for the evaluation of models addressing the issue of seabed segmentation. While some tiles could be described in a couple polygons, the ones depicting the hardest terrains with interlaced boundaries are comprised of dozens of polygons. The most challenging tile, and its one hundred polygons, is put aside from the Train set, for performances



(a) Ground truth map of the Unseen set.



(b) D4SC's predictions for a 87.2% pixel-wise accuracy.

Rocks
Flat Bottom
Small Sand Ripples
Large Sand Ripples

Figure 3: D4SC's predictions over the Unseen set (b) compared to the ground truth (a).

evaluation on unknown seabed configurations. This very tile comprises the Unseen set and its ground truth map is reported in Fig. 3(a).

4 Results

4.1 Training setup

The training of D4SC is performed within a Distributed Data Parallel (DDP) strategy with careful choice of training hyperparameters. The 2176 pixel coordinates of the Train dataset, producing as many randomly augmented 256×256 pixels patches every epoch, are split into training, validation and test set. To prevent from getting chance results by not training the CNN enough, an early stopping strategy with patience was employed to continue backpropagating the cross-entropy loss until no improvement over validation can be observed. That design choice enforces results to be more replicable. In the end, the training phase, which ranged from a couple of minutes to few hours, was fast enough to test different meta-architecture parameters and resulted in the final D4SC architecture.

4.2 Comparison with standard DL approaches

For deep learning based segmentation, prior arts applied domain adaptation with transfer learning or retrained from scratch models without tailoring their architecture neither to their data nor to their seabed characterisation task. Therefore, to compare our model, we reimplement the U-Net model[8], extended with batch normalisation layers, and we fine-tune a Deeplab[9] with a ResNet18 feature extractor trained on ImageNet, as standard transfer learning schemes. Despite U-Net's and Deeplab's costs, D4SC outperforms them while being a lot smaller, as reported in Table 1. Regardless of the small performance drop of D4SC over the Unseen set with its 87.2% reported accuracy and its predictions being slightly biased towards Rocks, resulting from the choice made in the annotation procedure, it succeeds in capturing homogeneous areas and most class boundaries, as depicted in Fig. 3(b), and forth in addressing the automatic semantic segmentation of the seabed over unknown configurations.

4.3 The effect of data augmentation

To ensure the added value of data augmentation over the complete learning pipeline, the performances of D4SC is evaluated with and without it, but also against the common data augmentation pipeline derived from computer vision. The results are reported in Table 2 and show that the novel data augmentation pipeline is able to extend the Train dataset and to ensure the generalisation

Table 1: Comparison of D4SC with U-Net[8] and Deeplab[9] in terms of pixel-wise Accuracy. The number of trainable parameters and Giga Floating-point Operations (GFLOPs) accounts for the simplicity of the model.

Approach	Trainable Parameters	GFLOPs	Accuracy	
			Train	Unseen
U-Net[8]	31M	138.34	0.956	0.864
Deeplab[9]	16M	11.06	0.964	0.833
D4SC	**958k**	**4.34**	**0.968**	**0.872**

Table 2: An evaluation of D4SC with, without augmentations and against common computer vision augmentations.

Approach	Accuracy	
	Train	Unseen
Without Augmentations	0.962	0.839
Common Augmentations	0.955	0.831
With our Augmentations	**0.968**	**0.872**

of the learned seabed semantic segmentation. Indeed, the misclassification results, mostly due to patches of limited dynamic or lacking recognisable patterns, are alleviated with our method.

5 Conclusion

The careful design of the novel learning pipeline of D4SC and its data augmentations are beneficial to seabed semantic segmentation by learning generic seabed patterns robust to noise and extending the dataset from a real-world survey. The good results over unseen configurations show that the proposed approach generalises well in low-label regime. Additionally, as our work did not make the most of the available unlabelled data, future research may investigate self-supervised or active learning methods to leverage even more the characterisation of the seabed.

References

[1] Maria Lianantonakis and Yvan R Pétillot. "Sidescan Sonar Segmentation Using Texture Descriptors and Active Contours". In: *IEEE JOE* 32 (2007), pp. 744–752.

[2] David P Williams. "Fast Target Detection in Synthetic Aperture Sonar Imagery: A New Algorithm and Large-Scale Performance Analysis". In: *IEEE JOE* 40 (1 2015), pp. 71–92. DOI: 10.1109/JOE.2013.2294532.

[3] David P Williams. "Fast Unsupervised Seafloor Characterization in Sonar Imagery Using Lacunarity". In: *IEEE TGRS* 53 (11 2015), pp. 6022–6034. DOI: 10.1109/TGRS.2015.2431322.

[4] Ze Liu et al. "Swin Transformer V2: Scaling Up Capacity and Resolution". In: *CVPR*. IEEE, June 2022, pp. 11999–12009. ISBN: 978-1-6654-6946-3. DOI: 10.1109/CVPR52688.2022.01170.

[5] Antoni Burguera and Francisco Bonin-Font. "On-Line Multi-Class Segmentation of Side-Scan Sonar Imagery Using an Autonomous Underwater Vehicle". In: *Journal of Marine Science and Engineering* 8 (8 2020). ISSN: 2077-1312. DOI: 10.3390/jmse8080557.

[6] Johnny Chen and Jason E. Summers. "Deep convolutional neural networks for semi-supervised learning from synthetic aperture sonar (SAS) images". In: *Proc. Mtgs. Acoust. 30*. 2017. DOI: 10.1121/2.0001018.

[7] Yung-Chen Sun, Isaac D. Gerg, and Vishal Monga. "Iterative, Deep Synthetic Aperture Sonar Image Segmentation". In: *IEEE TGRS* 60 (Sept. 2022), pp. 1–15. ISSN: 0196-2892. DOI: 10.1109/TGRS.2022.3162420.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Proc. MICCAI*. Springer International Publishing, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[9] Liang-Chieh Chen et al. "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation". In: *ECCV*. Feb. 2018, pp. 833–851. DOI: 10.1007/978-3-030-01234-2_49.

[10] Mark Sandler et al. "MobileNetV2: Inverted Residuals and Linear Bottlenecks". In: *CVPR*. IEEE, June 2018, pp. 4510–4520. ISBN: 978-1-5386-6420-9. DOI: 10.1109/CVPR.2018.00474.

[11] Mingxing Tan and Quoc V. Le. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks". In: *unpublished* (May 2019).

[12] David P Williams. "The Mondrian Detection Algorithm for Sonar Imagery". In: *IEEE TGRS* 56 (2 2018), pp. 1091–1102. DOI: 10.1109/TGRS.2017.2758808.