KELP WANTED: SEGMENTING KELP FORESTS

María Rodríguez, Candela Pelliza, Khizer Zakir, Leila Maritim

Universite Bretagne Sud (UBS), Vannes, France

ABSTRACT

Previous kelp mapping studies have often encountered limitations with traditional remote sensing approaches, especially due to the dynamic nature of kelp ecosystems. Deep learning semantic segmentation techniques offer a promising solution for the automatic detection of kelp. However, the effectiveness of these frameworks heavily depends on data quality, preprocessing approaches, the network architecture design and the choice of the loss function. In this study we explore a deep learning approach comparing SegNet and UNet-based architectures to detect kelp presence. We introduce a series of data-augmentation and data processing techniques. Our experimental findings highlight the effectiveness of different neural network architectures and loss functions in kelp segmentation that, combined with rule-based approaches, offer a boost in kelp detection accuracy.

Index Terms— Kelp mapping, semantic segmentation, deep learning

1. INTRODUCTION

Kelp forests are vital underwater ecosystems predominantly found in cold shallow waters. They are characterized by large brown fast-growing algae rooted in the seabed. As a foundational habitat for a myriad of sea organisms, these forests are the most diverse ecosystems in the ocean [3]. The two main species of canopy-forming kelps are giant kelps (Macrocystis pyifera) and bull kelp (Nereocystis leutkeana). Economically, kelp holds immense value, being harvested for various purposes such as food, fertilizers, and pharmaceuticals. It is estimated to generate an average revenue of $500 billion per year[1].

Despite their importance, kelp forests are facing significant challenges globally, becoming increasingly threatened by multiple drivers [1]. A combination of natural factors such as grazing pressure from urchins, wave action, storms, and rising temperatures, along with human-induced disturbances like shoreline development, pollution-induced ocean acidification, and destructive fishing practices, are contributing to an increasing decline in kelp forests development.

Detecting and tracking the presence, distribution, abundance, and health of kelp forests at different spatial scales is essential for effective conservation efforts. Optical remote sensing emerges as a powerful tool for quantifying and characterizing kelp, leveraging the high reflectance of kelp biomass in the Near Infrared compared to seawater. However, traditional spectral indices like the Normalized Difference Algae Index (NDAI), Normalized Difference Vegetation Index (NDVI), and Scaled Algae Index (SAI, while useful, have limitations due to atmospheric and environmental variations. These variations affect the density of the kelp canopy, subsequently influencing the strength of the signal detected by remote sensing sensors. To address these limitations, spectral unmixing approaches using hyperspectral imagery have shown promise in detecting bull kelp, particularly given their small bed sizes relative to spatial resolution[2]. Furthermore, drone imagery provides high-resolution data for visual counts, yet these methods are constrained by cost, time, and spatial extent[5].

Acknowledging the dynamic nature of kelp ecosystems and the variability in mapping challenges across locations, our study aims to explore adaptable methods to detect kelp presence or absence. Our study focuses on coastal areas along the Falklands/Malvinas Islands region, considering the typical growth pattern of kelps along rocky reefs in temperate and subpolar regions of the world[3]. To address this, we have implemented a deep-learning approach that facilitates automatic detection. We leverage the Segnet, and two UNet-based architectures for semantic segmentation to develop robust and versatile models for accurately detecting kelp across various environmental conditions.

In the following sections, we present our data as well as the pre-processing steps undertaken to enhance kelp features and eliminate noise. We then describe the deep learning approach taken together with the various experimental setups and rule-based post processing workflow. Finally, we present the results and conclusions.

2. METHODS AND MATERIALS

2.1. Data Description & pre-processing

This study utilizes the Kelp dataset provided by the Kelp Wanted Driven Data competition, consisting of two subsets. The first one comprises feature images obtained through remote sensing of the Falklands/Malvinas Islands, a region renowned for abundant kelp presence. These images are composed of seven bands: SWIR, NIR, Red, Green, and Blue spectral bands from Landsat 2, along with a Digital Elevation Model (DEM) sea-level derived from ASTER, and a cloud mask. All images maintain a spatial resolution of 30 meters and lack geolocation information. The second

subset contains binary label masks indicating kelp absence or presence pixel-wise, developed by citizen scientists via the Floating Forests platform.

Starting with the provided feature bands, we derived two additional subproducts that were integrated into the dataset as new bands: NDVI and Distance Map (DM). NDVI is a well-known vegetation index calculated using NIR and red bands, that has demonstrated utility in identifying kelp and other types of floating algae [3]. Distance Maps aim to include information about the distance of water pixels to the coastline, which was detected using Gaussian and canny edge detector filters to the DEM. They were generated by computing the Euclidean distance transform of each pixel in the background to the nearest coastal edge pixel. In instances where a landmass is absent, we set the distance to a predetermined constant value. This is aimed at mitigating anomalies in the distance maps.

Given the constraints of the competition, the original dataset comprised a training set consisting of 5637 feature images with corresponding labels, and a test set containing 1427 feature images without labels, all stored as 350x350 pixels TIFF files. Due to this configuration, we further divided the training set into train, validation, and test subsets for experimentation. Consequently, we worked with two distinct test sets: the Driven Data test set, comprising the original 1427 feature images without labels used for generating final predictions, and the Implemented test set, containing both features and label images for evaluating model performance.

Both the training and Driven Data test set datasets exhibit specific characteristics and data issues, as depicted in Fig. 1, which influenced most of our data processing and model development steps. The training dataset notably suffers from a significant class imbalance, with non-kelp pixels comprising 97.33% and kelp pixels only 0.67%. Additionally, both datasets contain a significant number of corrupted and missing value pixels, often exhibiting patterned artifacts due to image acquisition and preprocessing. The presence of cloud pixels, a well-known issue in remote sensing imagery, poses an additional challenge, with the provided cloud mask frequently misidentifying coastal areas as clouds. Moreover, variations in the percentage of water coverage across images complicate distance map calculation and kelp detection, particularly in images devoid of land pixels.

To address the mentioned data issues, we implemented several preprocessing steps. Initially, we normalized the data by calculating the standard deviation (σ) and mean statistics after excluding the lowest and highest 5% of data values. This normalization procedure helped prevent extreme values from influencing the calculation.

Next, to ensure a balanced distribution of kelp pixels across the training and validation subsets, we performed a stratified data split based on the distribution of kelp pixel percentage per image. By maintaining a proportional representation of kelp pixels in both subsets, this approach aimed to mitigate biases and enhance the models' generalization capability during training and evaluation.

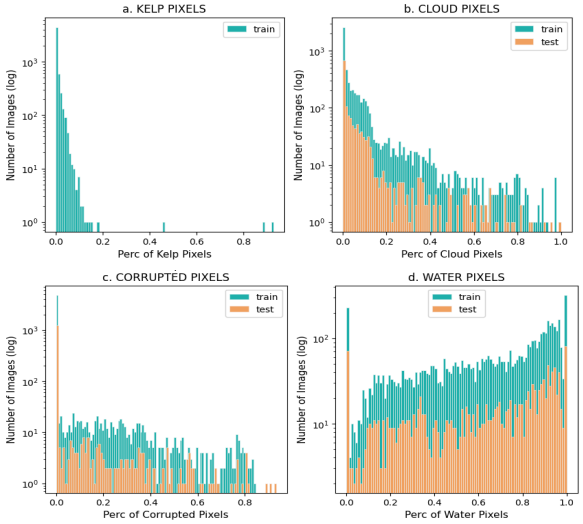

Fig. 1: Distribution of (a) kelp pixels, (b) cloud pixels, (c) corrupted pixels and (d) water pixels as percentage of the total pixels, across the images in train and Driven Data test sets.

Finally, we filtered the new training subset to include only images meeting specific criteria: at least 1 kelp pixel presence, less than 20% cloud cover, and less than 10% corrupted pixels. This filtering process helped improve the quality of the training data by excluding images with substantial cloud cover or pixel corruption.

2.2. Architectures

Our study delves into exploring the efficacy of three prominent Neural Network (NN) architectures for semantic segmentation tasks: SegNet, U-Net, and U-Net++. These architectures are well-suited for segmentation problems similar to the one addressed in this paper. They leverage deep learning techniques in learning complex patterns and relationships from data. In addition, the encoder-decoder structure with skip connections and nested architectures enables effective feature extraction and spatial context preservation, essential for segmentation tasks.

SegNet[5] employs a fully convolutional neural network (CNN) architecture for pixel-wise semantic segmentation. It utilizes an encoder-decoder framework but differs from U-Net in its use of max-pooling indices for upsampling during decoding. This approach enables efficient memory utilization and computational speed.

On the other hand, U-Net[4]is characterized by a symmetric encoder-decoder architecture, which consists of a contracting path to capture context and a symmetric expanding path for precise localization. Its skip connections aid in preserving spatial information, making it suitable for tasks where precise delineation is crucial.

A refinement of the original U-Net architecture, U-Net++[7] was proposed to address limitations in capturing hierarchical features. It incorporates dense skip connections and nested architectures, facilitating the extraction of multi-scale features and enhancing segmentation accuracy, particularly in complex scenes.

2.3. Loss Functions and Metrics

The Dice Loss, derived from the dice coefficient, finds wide application in segmentation tasks as it quantifies the overlap between predicted and ground truth masks. Furthermore, combining Cross-Entropy (CE) or Balanced Cross-Entropy (BCE) loss with the Dice Loss penalizes misclassifications and addresses the trade-off between precision and recall in segmentation tasks[7]. By assigning weights to different classes or regions of interest, it enables more flexible optimization based on the importance of specific regions and addresses class imbalance issues.

Considering the class imbalance in our data, we conducted various experiments using Dice Loss, CE+Dice, and BCE+Dice.

$$Dice(A,\ B) = \frac{2||A \cap B||}{||A|| + ||B||}\ ,\ IoU(A,B) = \frac{||A \cap B||}{||A \cup B||} \text{ — eq (1)}$$

We assess the segmentation results using the recommended dice coefficient and an additional metric, IoU, eq (1). Both metrics range between 0 (indicating no overlap) and 1 (indicating a perfect match between predicted and ground truth masks). While the dice coefficient provides an overall measure of overlap, IoU emphasizes individual instances of misclassification more, potentially magnifying the impact of outliers on performance evaluation.

2.4. Post-Processing

In post-processing, the predicted outcomes were filtered using a simple mask for the presence of corrupt pixels and DEM > 0 based on statistical observations. When either one of the conditions is satisfied, the pixel is re-classified as non-kelp. This rule-based step helps enhance the accuracy of the segmentation outputs.

3. RESULTS

3.1. Experimental Settings

As a baseline for our experiments, we initially utilized the SegNet architecture. The training was conducted using the spectral bands (SBs) (SWIR, NIR, Red, Green, Blue) and DMs. The setup included a batch size of 8, the use of Adam optimizer, Dice Loss function, and a learning rate of 10^{-3} over 20 epochs.

Building on this baseline, we optimized the model hyperparameters. Specifically, we experimented with batch sizes of 8 and 16 and varied the learning rate logarithmically from 10^{-2} to 10^{-4}. To ensure the model convergence during the training phase, we tested 20, 50, and 80 total epochs. Additionally, we integrated techniques such as batch normalization, weight decay (evaluating values of 10^{-7} and 10^{-8}) and early stopping with patience levels of 5 and 10. This hyperparameter tuning phase led to an optimized configuration of a 16 batch size, a learning rate of 10^{-4}, 80 epochs, a weight decay of 10^{-7}, and an early stopping with a patience of 10.

Subsequently, we explored the potential of architectures with a larger parameter set, such as U-Net, starting with the optimized settings derived from our SegNet experiments. This included a comparative analysis with the U-Net framework to determine how different loss functions and spectral band combinations influenced segmentation results. In terms of spectral band setups, we explored a range of combinations, such as SBs, DMs, DEM, and the NDVI.

Following the insights gained from these experiments, we trained the final SegNet and UNet models using the optimal loss function and spectral band combination identified and including the post-processing techniques outlined on Section 2.4. Finally, we utilized the U-Net++ architecture, known for its increased parameter count, to potentially enhance segmentation metrics. This model was trained under the same conditions as those used for the SegNet and U-Net models.

All experiments were conducted within the PyTorch framework. The SegNet and U-Net models were trained on a Tesla P100 GPU with 16GB of memory, while the U-Net++ architecture required an A6000 GPU with 48GB due to its larger size.

3.2. Experimental Results

Table 1 shows the results of evaluating the three distinct loss functions. In our implemented test set, the CE + Dice Loss function produced the best Dice Coefficient and IoU scores. Conversely, the Dice Loss outperformed others in the Driven Data test set, with both CE+Dice Loss and BCE+Dice Loss recording a coefficient under 0.5. Further analysis of kelp image predictions showed that in the Driven Data test set, images with striping acquisition errors were

inaccurately identified as kelp by these last two loss functions.

Loss function	Driven Data test set	Implemented test set	
	Dice	Dice	IoU
Dice	**0.6946**	0.6700	0.5000
CE + Dice Loss	0.1560	**0.6790**	**0.5150**
BCE + Dice	0.3786	0.6590	0.4910

Table. 1: Comparative Evaluation of Segmentation Performance using distinct Loss Functions

For the band combination configurations, the inclusion of SBs, DMs, and NDVI resulted in the highest Dice Coefficient and IoU in our implemented test set, as shown in Table 2. In contrast, for the Driven Data test, the optimal performance was achieved using only SBs and DMs. Introducing NDVI slightly decreased the Dice score from 0.6946 to 0.6942.

Bands	Driven Data test set	Implemented test set	
	Dice	Dice	IoU
SB	**x**	0.6625	0.4953
SB + DM	**0.6946**	0.6700	0.5000
SB + DM + NDVI	0.6942	**0.6819**	**0.5174**
SB + DM + NDVI+ DEM	0.6877	0.6612	0.4938

Table. 2: Comparative results of Segmentation Performance across different Band Configurations

Our final model evaluations showed that for the Implemented test set, models with higher complexity and more trainable parameters, such as Unet ++, yielded improved performance metrics, as outlined in Table 3. This pattern, however, was not replicated in the Driven Data test set, where the Unet architecture, despite its lower parameter count compared to Unet ++, achieved superior results.

Architecture	# Trainable Parameters	Driven Data test set	Implemented test set	
		Dice	Dice	IoU
Segnet	29,445,315	0.6723	0.6631	0.4960
Unet	31,392,449	**0.6976**	0.6879	0.5243
Unet ++	36,631,361	0.6945	**0.6929**	**0.5301**

Table. 3: Final Performance Metrics of various architectures

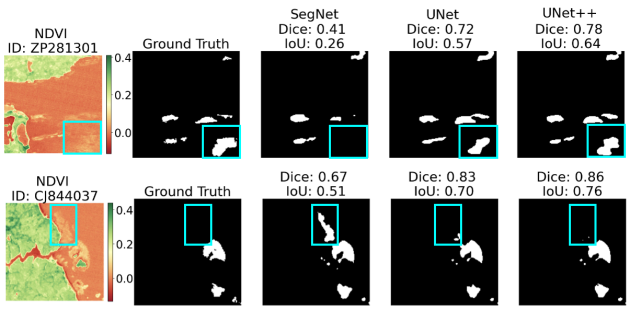

Fig. 2: Qualitative assessment of prediction outputs made by the final models

Finally, Fig. 2 presents a comparison between the ground truth data and the prediction outputs from the final models. It shows that the SegNet model struggles with accurately classifying kelp pixels, resulting in false negatives, whereas the U-Net models successfully detect them. Additionally, SegNet misclassified some water pixels as kelp, leading to false positives.

4. CONCLUSION

Our research delves into employing semantics segmentation deep learning approaches for kelp detection, showcasing their efficacy in addressing the challenges of traditional remote sensing methods in kelp segmentation. We found that data preprocessing, the selection of the model architecture, and parameter settings significantly impacts segmentation accuracy, particularly in scenarios, where class imbalance is prevalent. Moreover, our study shows that the inclusion of DMs, together with rule-based post processing workflows can effectively increase the model accuracy. We recognized that, even after the implemented processing steps, corrupted images are still an important issue leading to kelp misclassification. Prospective studies should dig into possible approaches to overcome data inaccuracies and corrupted images.

5. REPRODUCIBILITY

The processed data and developed code is available at: https://github.com/mariarodriguezn/KelpSemanticSegmentation (Accessed on February 22, 2024)

6. REFERENCES

[1] A. M. Eger *et al.*, "The value of ecosystem services in global marine kelp forests," *Nat. Commun.*, vol. 14, no. 1, p. 1894, Apr. 2023, doi: 10.1038/s41467-023-37385-0.

[2]Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561, 5.

[3] H. Teagle, S. J. Hawkins, P. J. Moore, and D. A. Smale, "The role of kelp species as biogenic habitat formers in coastal marine ecosystems," *J. Exp. Mar. Biol. Ecol.*, vol. 492, pp. 81–98, Jul. 2017, doi: 10.1016/j.jembe.2017.01.017.

[4] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Lecture Notes in Computer Science, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

[5] S. B. Schroeder, C. Dupont, L. Boyer, F. Juanes, and M. Costa, "Passive remote sensing technology for mapping bull kelp (Nereocystis luetkeana): A review of techniques and regional case study," *Glob. Ecol. Conserv.*, vol. 19, p. e00683, Jul. 2019, doi: 10.1016/j.gecco.2019.e00683.

[6] Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Jorge Cardoso, M. (2017). Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 240–248. https://doi.org/10.1007/978-3-319-67558-9_28

[7] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018a). UNet++: A nested U-Net Architecture for Medical Image segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 3–11. https://doi.org/10.1007/978-3-030-00889-5_1