

Project Team 39 Proposal

1. Team Members:

- Nadira Amadou Mahamane
- Oluwatofunmi Sodimu
- Thanapol Tantagunninat

2. Problem Definition: What are you trying to solve or do? Why is this a problem that is useful or interesting to study or experiment with? (2-3 bullet points)

- The goal of our project is to create a model that can detect the presence of kelp forests in coastal satellite images. Existing methods, like citizen science classification, are labor-intensive and slow, requiring a more efficient approach for comprehensive global monitoring.
- This problem is interesting because the ability to map and monitor kelp forests will enable us to preserve and protect them from climate change, overfishing, unsustainable harvesting practices and other threats. Mapping and monitoring of kelp forests, crucial underwater habitats, are essential due to their significance in coastal marine ecosystems and their economic value.

3. Project Idea: What do you plan on accomplishing in your project scope? How will your idea solve the problem proposed in your problem definition? (2-3 bullet points)

- The goal is to develop a binary semantic segmentation algorithm that can detect the presence of kelp canopies using Landsat satellite images and labels from the 'Driven Data' website.
- This project is a step towards the development of models that can accurately map and monitor kelp forests which are essential to many ecosystems and species.
- The success of this project also provides a scalable and adaptable solution for global application.

4. Related Works: Briefly describe what work has been done in this area and how they currently approach the problem. List any research papers, existing codebases, or miscellaneous articles/websites that you found that relate to your work. (minimum 5 references; at least 3 should be research papers)

- ⊄ Previous research has suggested the use of a Mask R-CNN (mask region-based convolutional neural network) to detect giant kelp forests along the coastlines of Southern California and Baja California using satellite imagery. The authors were able to achieve high detection accuracy while minimizing overprediction by

optimizing the mask R-CNN model through hyper parameterization. This study highlights the potential of Mask R-CNN as a valuable tool for marine ecological monitoring, which can help in informed decision-making processes related to biodiversity conservation and management.

Marquez, L., Fragkopoulou, E., Cavanaugh, K.C. *et al.* Artificial intelligence convolutional neural networks map giant kelp forests from satellite imagery. *Sci Rep* **12**, 22196 (2022).
<https://doi.org/10.1038/s41598-022-26439-w>

- ∄ The paper "Automatic Hierarchical Classification of Kelps using Deep Residual Learning" presents an approach for the automatic hierarchical classification of kelps using deep residual learning for image recognition. The study focuses on the classification of kelps using deep learning techniques and computer vision algorithms.

Mahmood A, Ospina AG, Bennamoun M, An S, Sohel F, Boussaid F, Hovey R, Fisher RB, Kendrick GA. Automatic Hierarchical Classification of Kelps Using Deep Residual Features. *Sensors*. 2020; 20(2):447.
<https://doi.org/10.3390/s20020447>

- ∄ The paper entitled "Automated Satellite Remote Sensing of Giant Kelp at the Falkland Islands (Islas Malvinas)" compares two approaches for automating the detection of giant kelp in satellite datasets. The study uses crowd-sourced classification and a spectral approach that includes decision trees and spectral unmixing. According to the paper, the canopy area of giant kelp is influenced by environmental factors such as nitrate content, although no long-term trends have been observed. The results highlight the importance of nitrate variability in controlling the giant kelp canopy area in the FLK region. The study underscores the significance of monitoring giant kelp populations and demonstrates the usefulness of remote sensing techniques for such assessments, especially in vast geographical areas.

Houskeeper HF, Rosenthal IS, Cavanaugh KC, et al. Automated satellite remote sensing of giant kelp at the Falkland Islands (Islas Malvinas). *PLoS One*. 2022;17(1):e0257933. Published 2022 Jan 6.
doi:10.1371/journal.pone.0257933

- ∄ The paper "Submerged Kelp Detection with Hyperspectral Data" describes a fully automated simple feature detection processor to detect the presence of kelp in submerged habitats using hyperspectral data. It compares different detection methods and discusses the validation of classification results

Uhl F, Bartsch I, Oppelt N. Submerged Kelp Detection with Hyperspectral Data. *Remote Sensing*. 2016; 8(6):487. <https://doi.org/10.3390/rs8060487>

- ∄ The article "An Automated Method for Mapping Giant Kelp Canopy Dynamics from UAV" presents a method for mapping giant kelp canopy dynamics using

unmanned aerial vehicles (UAVs). The study focuses on the development of a rapid survey method for determining canopy biomass, with specific emphasis on kelp patch-specific characteristics that limit detection capability. The research is significant due to the dominance of kelp forests in autotrophic biomass and primary productivity in shallow temperate and Arctic rocky areas. The article provides valuable insights into the use of UAVs for mapping kelp canopy dynamics, which is essential for understanding and monitoring kelp ecosystems.

Doubkova, M., & Kudela, R. (2020). An Automated Method for Mapping Giant Kelp Canopy Dynamics from UAV. *Frontiers in Environmental Science*, 8, 587354. <https://doi.org/10.3389/fenvs.2020.587354>

5. **Datasets:** How will you benchmark your approach or theory? Describe and link each dataset you will use in your project and explain why you selected the dataset. (min 1 dataset)

- This project was inspired by a competition on the driven data website. So, we intend to use the performance metric listed on the competition page – Dice Coefficient – to determine our segmentation accuracy and assess our model. In addition, we will also utilize other evaluation metrics for segmentation tasks such as Accuracy, Precision, Recall, F1, IoU, and Area under ROC curve.
- Dataset: <https://www.drivendata.org/competitions/255/kelp-forest-segmentation/data/>
- We selected the dataset in the link above because:
 - It consists of images with 7 channels, including Short-wave IR, Near-IR, RGB, Cloud Mask, and Land Mask that could be utilized for the benefit of segmenting the Giant Kelp.
 - The training set includes 5,635 TIF images/label pairs, and the test set includes 1,426 TIF images. Since the ground truth label for the test set has not been released, we will split the given train image set which we think is sufficient for training and testing of the model.
 - The dataset was curated by a group of scientists from UMass Boston and Woods Hole Oceanographic Institution. Thus, we expect that the data will be accurate and reliable.

6. **Proposed Experiments:** What experiments will your team perform in your project? Roughly describe what ideas you will try. (2-3 bullet points)

- Data Processing/Augmentation: train the model with different image channels and combination of image channels, removing the cloud and/or land for the model, adding filters, changing image contrast, and measure the performance.
- Binary Semantic Segmentation using deep learning algorithms such as convolutional neural networks (CNN).

- Evaluate model robustness by introducing simulated data issues such as noise, cloud cover, and variations in lighting conditions to assess real-world applicability.

7. **Compute Resources:** What compute resources (GPUs, CPUs, TPUs) does your team plan to use? The type of compute resources can limit the scope of your project, so we want teams to pick appropriate projects for the resources they have accessible.

- PACE ICE Cluster
- Google Colab
- Onboard GPU