Coral Grief: Machine Learning on Crowd-sourced Data to Highlight an Ecological Crisis

Narayan, Rithika (School: Elwood John Glenn High School)

Triggered largely by the warming and pollution of oceans, corals are experiencing bleaching and a variety of diseases caused by the spread of bacteria, fungi, and viruses. Identification of bleached and/or diseased corals enables implementation of measures to halt or retard the same. Benthic cover analysis as a standalone measure of reef health is insufficient for identification of coral bleaching and/or disease. Proposed herein is a solution that couples machine learning with crowd-sourced data - images from government archives, citizen science projects, and personal images collected by tourists - to build a model capable of identifying healthy, bleached, and/or diseased coral. The student researcher collected hundreds of images of corals with various health conditions from open sources such as the National Oceanic and Atmospheric Administration's records and the XL Catlin Seaview Survey and annotated these images using the image annotation platform Labelbox in order to highlight the regions of interest: healthy, bleached, black band disease, dark spot disease, white syndrome, or yellow band disease. These annotations were then used to build, train, and validate a Python-based model, adapted from an open source Mask R-CNN (region-based convolutional neural network) algorithm, within an Amazon Web Services EC2 remote computer. Use of the model on a test set of coral images yields over 85% accuracy in distinguishing healthy versus unhealthy coral. This machine learning-based model has the potential to rapidly analyze a large and growing database of images to identify coral bleaching and/or coral disease around the world, thereby enabling effective allocation of resources for preservation of our marine ecosystem.

Awards Won: